

EXPANDING THURSTON MAPS

MARIO BONK AND DANIEL MEYER

ABSTRACT. We study the dynamics of Thurston maps under iteration. These are branched covering maps f of 2-spheres S^2 with a finite set $\text{post}(f)$ of postcritical points. We also assume that the maps are expanding in a suitable sense. Every expanding Thurston map is a factor of a shift operator. This link to symbolic dynamics suggests our ultimate goal of finding a combinatorial description of the dynamics of an expanding Thurston map in terms of finite data. Relevant for this problem are existence and uniqueness results for f -invariant Jordan curves $\mathcal{C} \subset S^2$ containing the set $\text{post}(f)$. For every sufficiently high iterate f^n of an expanding Thurston map such an invariant Jordan curve always exists. If the sphere S^2 is equipped with a “visual” metric d adapted to the dynamics of f , then an f -invariant Jordan curve \mathcal{C} with $\text{post}(f) \subset \mathcal{C}$ is a quasicircle. The geometry of the metric space (S^2, d) encodes many dynamical properties of f . For example, $f: S^2 \rightarrow S^2$ is topologically conjugate to a rational map if and only if (S^2, d) is quasimetrically equivalent to the Riemann sphere $\widehat{\mathbb{C}}$.

Establishing a framework for proving these and other results for expanding Thurston maps is the main purpose of this work.

CONTENTS

List of Figures	2
1. Introduction	3
1.1. Results	4
1.2. Lattès maps	9
1.3. Related work	12
1.4. Outline of the paper and main ideas	12
1.5. Examples of Thurston maps	16
2. Notation	17

Date: September 19th, 2010.

Key words and phrases. Expanding Thurston map, postcritically-finite map, visual metric, invariant curve, Markov partition.

M.B. was supported by NSF grants DMS 024421, DMS 0456940, and DMS 0652915.

D.M. was supported by an NSF postdoctoral fellowship and the Academy of Finland, projects SA-134757 and SA-118634.

3. Thurston maps	18
4. Cell decompositions	21
5. Cell decompositions of 2-spheres	28
6. Cell decompositions induced by Thurston maps	41
7. Flowers	47
8. Expansion and visual metrics	55
9. Symbolic dynamics	69
10. Isotopies	77
10.1. Equivalent expanding Thurston maps are topologically conjugate	78
10.2. Isotopies of Jordan curves	84
11. Thurston maps with invariant curves	89
12. Two-tile subdivision rules	98
12.1. Examples of two-tile subdivision rules	115
13. Combinatorially expanding Thurston maps	124
14. Auxiliary results on graphs	146
15. Invariant curves	154
15.1. Existence and uniqueness of invariant curves	157
15.2. Iterative construction of invariant curves	166
16. Invariant curves are quasicircles	178
17. Periodic critical points	186
18. The combinatorial expansion factor	189
19. Rational Thurston maps	200
20. The measure of maximal entropy	211
21. Concluding remarks and open problems	230
References	236
Index	240

LIST OF FIGURES

1 The Lattès map g .	10
2 Cycle of a vertex v .	31
3 J is not isotopic to S^1 rel. $\{1, \mathbf{i}, -1, -\mathbf{i}\}$.	85
4 The proof of Lemma 11.3.	93
5 The two-tile subdivision rule for $z^2 - 1$.	116
6 Tiles of order 7 for Example 12.10.	116
7 The barycentric subdivision rule.	117
8 Tiles of order 4 for the barycentric subdivision rule.	118
9 The subdivision rule for Example 12.12.	119
10 The 2-by-3 subdivision rule.	120

11	The subdivision rule realized by $1 - 2/z^2$.	122
12	Adding a flap.	123
13	The subdivision rule realized by f_6 .	123
14	Tiles of order 4 of the map f_6 .	124
15	The map f is not combinatorially expanding, but equivalent to the expanding map g .	145
16	Constructing a path through a, b, p .	148
17	Construction of the curve \mathcal{C}' .	153
18	The invariant curve for Example 15.1.	155
19	Invariant curves for the Lattès map g .	159
20	No invariant Jordan curve $\tilde{\mathcal{C}} \supset \text{post}$.	161
21	Iterative construction of an invariant curve.	168
22	Iterative construction by replacing edges.	175
23	Example where $\tilde{\mathcal{C}}$ is not a Jordan curve.	177
24	A non-trivial rectifiable invariant Jordan curve.	177
25	Replacing k -tiles by $(k - 1)$ -tiles.	194

1. INTRODUCTION

Let $f: \hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$ be a rational map on the Riemann sphere $\hat{\mathbb{C}}$ of degree ≥ 2 . As usual we call a point $p \in \hat{\mathbb{C}}$ a *critical point* of f if near p the map f is not a local homeomorphism. A *postcritical point* is any point obtained as an image of a critical point under forward iteration of f . So if we denote by $\text{crit}(f)$ the set of critical points of f and by f^n the n -th iterate of f , then the set of postcritical points of f is given by

$$\text{post}(f) := \bigcup_{n \geq 1} \{f^n(c) : c \in \text{crit}(f)\}.$$

It is a fundamental fact in complex dynamics that much information on the dynamics of f can be deduced from the structure of the orbits of critical points. A very strong assumption in this respect is that each such orbit is finite, i.e., that the set $\text{post}(f)$ is a finite set. In this case the map f is called *postcritically-finite*.

A characterization of such maps is due to Thurston. The framework for his investigations was the setting of branched covering maps of 2-spheres. These are continuous maps $f: S^2 \rightarrow S^2$ on an oriented 2-sphere S^2 that near each point can be written as $z \mapsto z^d$ after suitable orientation-preserving coordinate changes in source and target. For

such maps the sets of critical and postcritical points can be defined in the same way as for rational maps on $\widehat{\mathbb{C}}$. A *Thurston map* is a branched covering map $f: S^2 \rightarrow S^2$ with a finite set of postcritical points that is not a homeomorphism. In this paper we study Thurston maps that are expanding in an appropriate sense. By definition this means that there exists a Jordan curve \mathcal{C} in S^2 containing the set of postcritical points of f such that the complementary components of $f^{-n}(\mathcal{C})$ become uniformly small as $n \rightarrow \infty$. For rational Thurston maps this is satisfied if and only if f has no periodic critical points. It is also equivalent to requiring that the Julia set $\mathcal{J}(f)$ of f is the whole Riemann sphere (see Proposition 19.1). Note that in contrast to the rational case, in general an expanding Thurston map may have periodic critical points; see Example 12.11.

1.1. Results. One of our goals is to obtain a description of the dynamics of an expanding Thurston map in terms of finite combinatorial data. A very general setting which allows one to address this question is the recently developed theory of self-similar group actions (see [Ne], in particular Sect. 6). Our approach is more concrete and based on the existence of f -invariant Jordan curves $\mathcal{C} \subset S^2$, i.e., Jordan curves with $f(\mathcal{C}) \subset \mathcal{C}$. We also require that $\text{post}(f) \subset \mathcal{C}$, because under this condition the preimages $f^{-n}(\mathcal{C})$ of \mathcal{C} under the iterates f^n of f form the 1-skeleton of a cell decomposition of S^2 that allows one to recover f^n as a “cellular map” in essentially combinatorial terms (see the discussion below). Some of our main results are about existence and uniqueness of such Jordan curves \mathcal{C} .

For rational Thurston maps we have the following statement.

Theorem 1.1. *Let $f: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ be a rational Thurston map with Julia set $\mathcal{J}(f) = \widehat{\mathbb{C}}$. Then for each sufficiently large $n \in \mathbb{N}$ there exists a quasicircle $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\text{post}(f) \subset \mathcal{C}$ and $f^n(\mathcal{C}) \subset \mathcal{C}$.*

If a curve \mathcal{C} is invariant for some iterate f^n , then one cannot expect it to be invariant for some other iterate f^k unless k is a multiple of n (see Remark 15.10). So in general, the curve \mathcal{C} in the previous theorem (also in Theorem 1.2 below) will depend on n . For the definition of a quasicircle see Section 16. Theorem 1.1 is a special case of a more general fact.

Theorem 1.2. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$. Then for each sufficiently large $n \in \mathbb{N}$ there exists a Jordan curve $\tilde{\mathcal{C}}$ that is invariant for f^n and isotopic to \mathcal{C} rel. $\text{post}(f)$.*

See Section 3 for a discussion of isotopies and related terminology. Since $\tilde{\mathcal{C}}$ is isotopic to \mathcal{C} rel. $\text{post}(f)$, it will also contain the set $\text{post}(f)$. The curve $\tilde{\mathcal{C}}$ is actually a quasicircle if S^2 is equipped with a suitable metric (see Theorem 1.8 below).

An obvious question is whether one can always find a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$ that is invariant under the map f itself, and hence invariant under all iterates f^n . It turns out that this is not true in general (Example 15.5), but one can give a necessary and sufficient condition for the existence of such an invariant curve.

Theorem 1.3 (Existence of invariant curves). *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map. Then the following conditions are equivalent:*

- (i) *There exists an f -invariant Jordan curve $\tilde{\mathcal{C}} \subset S^2$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$.*
- (ii) *There exist Jordan curves $\mathcal{C}, \mathcal{C}' \subset S^2$ with $\text{post}(f) \subset \mathcal{C}, \mathcal{C}'$ and $\mathcal{C}' \subset f^{-1}(\mathcal{C})$, and an isotopy $H: S^2 \times I \rightarrow S^2$ rel. $\text{post}(f)$ with $H_0 = \text{id}_{S^2}$ and $H_1(\mathcal{C}) = \mathcal{C}'$ such that the map*

$$\hat{f} := H_1 \circ f \text{ is combinatorially expanding for } \mathcal{C}'.$$

Moreover, if (ii) is true, then there exists an f -invariant Jordan curve $\tilde{\mathcal{C}} \subset S^2$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$ that is isotopic to \mathcal{C} rel. $\text{post}(f)$ and isotopic to \mathcal{C}' rel. $f^{-1}(\text{post}(f))$.

See Definition 11.4 for the notion of an combinatorially expanding Thurston map (Definition 7.6 and (7.4) are also relevant here).

The condition of combinatorial expansion in (ii) is easy to check in general (see Remark 15.7 (a), Proposition 15.13, and the examples discussed in Section 15). One can actually give a criterion for the existence of an invariant curve in a given isotopy class rel. $\text{post}(f)$ or rel. $f^{-1}(\text{post}(f))$ (see Remark 15.7 (c) and Proposition 15.8). Moreover, if an f -invariant Jordan curve \mathcal{C} exists, then it is the Hausdorff limit of a sequence of Jordan curves \mathcal{C}^n that can be obtained from a simple iterative procedure (Proposition 15.14).

Our existence results are complemented by the following uniqueness statement for invariant Jordan curves.

Theorem 1.4. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map and \mathcal{C} and \mathcal{C}' be f -invariant Jordan curves in S^2 that both contain the set $\text{post}(f)$. Then $\mathcal{C} = \mathcal{C}'$ if and only if \mathcal{C} and \mathcal{C}' are isotopic rel. $f^{-1}(\text{post}(f))$.*

As a consequence one can prove that if $\#\text{post}(f) = 3$, then there are at most finitely many f -invariant Jordan curves $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$

\mathcal{C} (Corollary 15.3). In the general case, there are at most finitely many such curves \mathcal{C} in a given isotopy class rel. $\text{post}(f)$ (Corollary 15.2). In general, a Thurston map may have infinitely many such invariant curves \mathcal{C} (Example 15.4).

Expanding Thurston maps are abundant and include specific maps on $\widehat{\mathbb{C}}$ such as $f(z) = 1 - 2/z^2$ or $f(z) = 1 + (i - 1)/z^4$. More examples can be found in Section 12.1. A large class of well-understood Thurston maps are *Lattès maps*. These are rational maps obtained as quotients of conformal torus endomorphisms (note that the terminology is not uniform and some authors use the term Lattès map with a slightly different meaning). We will discuss an explicit Lattès map in detail below.

A general method for producing Thurston maps is given by Proposition 12.5 in combination with Corollary 13.18; conversely, as follows from Theorem 1.2, at least some iterate of every expanding Thurston map can be obtained from this construction.

If $f: S^2 \rightarrow S^2$ is an expanding Thurston map and $\mathcal{C} \subset S^2$ a Jordan curve with $\text{post}(f) \subset \mathcal{C}$, then for each $n \in \mathbb{N}_0$ one can define an associated cell decomposition $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ of S^2 (Section 6). Its 0-skeleton is the set $f^{-n}(\text{post}(f))$, and its 1-skeleton the set of $f^{-n}(\mathcal{C})$ (see Section 4 for the terminology). In general, these cell decompositions \mathcal{D}^n are not compatible for different levels n , but if the curve \mathcal{C} is f -invariant, then \mathcal{D}^{n+1} is a refinement of \mathcal{D}^n for each $n \in \mathbb{N}$, and the pair $(\mathcal{D}^1, \mathcal{D}^0)$ gives a *cellular Markov partition* (Definition 4.8) for f that determines the combinatorics of all cell decompositions \mathcal{D}^n . This cellular Markov partition for f is of a particular type, namely coming from a *two-tile subdivision rule* (see Section 12).

The main consequence of Theorem 1.2 is that we get such a two-tile subdivision rule for some iterate $F = f^n$ of every expanding Thurston map. This essentially allows one to describe the map F in terms of finite combinatorial data.

Corollary 1.5. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map. Then for each sufficiently large n there exists a two-tile subdivision rule that is realized by $F = f^n$.*

If we allow more general cellular Markov partitions, it seems very likely that not only an iterate of f , but f itself allows a cellular Markov partition.

Conjecture. *Every expanding Thurston map admits a cellular Markov partition.*

We believe that even though we were not able to settle the conjecture, the methods established in this paper will be useful for answering this question.

We will see that an expanding Thurston map $f: S^2 \rightarrow S^2$ induces a natural class of metrics on S^2 that we call *visual metrics* (see Section 8). Each visual metric d has an associated *expansion factor* $\Lambda > 1$ and is characterized by the geometric property that for cells σ, τ in the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ we have $\text{diam}_d(\sigma) \asymp \Lambda^{-n}$ if σ has dimension ≥ 1 and $\text{dist}_d(\sigma, \tau) \gtrsim \Lambda^{-n}$ if $\sigma \cap \tau = \emptyset$ (Lemma 8.10). Any two visual metrics are *snowflake equivalent*, and the class of visual metrics for f and any iterate of f are the same.

The properties of visual metrics are essential in proving the following statement (see Section 9 for the relevant definitions).

Theorem 1.6. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map. Then f is a factor of the left-shift $\Sigma: J^\omega \rightarrow J^\omega$ on the space J^ω of all sequences in a finite set J of cardinality $\#J = \deg(f)$.*

An immediate consequence of this theorem and its proof is the fact that the periodic points of an expanding Thurston map $f: S^2 \rightarrow S^2$ are dense in S^2 (Corollary 9.3).

The proof of Theorem 1.6 is a simple adaption of the proof of a similar statement in [Ka, Thm. 3.4]. The basic idea seems to go back to [Jo] (see also [Prz]).

Our choice of the term “visual metric” is motivated by the close relation of this concept to the notion of a visual metric on the boundary of a Gromov hyperbolic space. For each expanding Thurston map $f: S^2 \rightarrow S^2$ one can construct a Gromov hyperbolic graph \mathcal{G} whose boundary at infinity $\partial_\infty \mathcal{G}$ can be identified with S^2 so that the class of visual metrics on $\partial_\infty \mathcal{G}$ in the sense of Gromov hyperbolic spaces is identical to the class of visual metrics for f in our sense (see Remark 8.8 for more explanation). In this paper we will not pursue this point of view further though (see [HP09] for an exposition of similar ideas).

It is possible to describe the range of possible expansion factors of visual metrics for an expanding Thurston map f . If d a visual metric with expansion factor $\Lambda > 1$, then a “snow-flaking” d^α with $\alpha \in (0, 1)$ results in a visual metric with the smaller expansion factor Λ^α . So the relevant problem is to find the supremum of all such expansion factors. This supremum is given by a *combinatorial expansion factor* $\Lambda_0(f)$ that can in principle be computed from combinatorial data (see Proposition 18.2 for the definition of $\Lambda_0(f)$).

Theorem 1.7. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map with combinatorial expansion factor $\Lambda_0(f)$.*

- (i) If Λ is the expansion factor of a visual metric for f , then $1 < \Lambda \leq \Lambda_0(f)$.
- (ii) Conversely, if $1 < \Lambda < \Lambda_0(f)$, then there exists a visual metric d for f with expansion factor Λ . Moreover, the visual metric d can be chosen to have the following additional property:
For every $x \in S^2$ there exists a neighborhood U_x of x such that

$$(1.1) \quad d(f(x), f(y)) = \Lambda d(x, y) \text{ for all } y \in U_x.$$

In particular, if f is an expanding Thurston map, then we can always find a visual metric d so that f scales the metric d by a constant factor at each point. The example of the Lattès map $g: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ discussed below illustrates this statement. If we equip $\widehat{\mathbb{C}}$ with a suitable visual metric for g (a flat orbifold metric with four conical singularities), then g behaves like a piecewise similarity map, where distances are scaled by the factor $\Lambda = 2$.

The combinatorial expansion factor $\Lambda_0(f)$ is invariant under topological conjugacy (Proposition 18.3) and well-behaved under iteration (see (18.3)).

The invariant curve \mathcal{C} in Theorem 1.2 equipped with (the restriction of) a visual metric is a quasicircle. This follows from the following general fact applied to the map f^n .

Theorem 1.8. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$. If \mathcal{C} is f -invariant, then \mathcal{C} equipped with (the restriction of) a visual metric for f is a quasicircle.*

Properties of f are encoded in the geometry of (S^2, d) , where d is a visual metric. For example, (S^2, d) is a *doubling metric space* if and only if f has no periodic critical points (see Section 17). One can also recognize when f is topologically conjugate to a rational map.

Theorem 1.9. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map, and d a visual metric for f . Then (S^2, d) is quasisymmetrically equivalent to $\widehat{\mathbb{C}}$ if and only if f is topologically conjugate to a rational map.*

Here $\widehat{\mathbb{C}}$ is equipped with the chordal metric σ . For the definition of quasisymmetric maps see Section 16.

If $f: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ is an expanding Thurston map and rational, then the last theorem implies that the chordal metric σ on $\widehat{\mathbb{C}}$ is quasisymmetrically equivalent to a visual metric; in general, σ is not a visual metric itself.

Thurston studied the question when a given Thurston map is represented by a conformal dynamical system from a slightly different

viewpoint. He asked when a Thurston map $f: S^2 \rightarrow S^2$ is in a suitable sense *equivalent* (see Definition 3.3) to a rational map and obtained a necessary and sufficient condition [DH]. For expanding Thurston maps his notion of equivalence actually means the same as topological conjugacy of the maps (Theorem 10.4).

The proof of Theorem 1.9 does not use Thurston's theorem mentioned above. Indeed, none of our results relies on this fact. Thus, our methods possibly provide a different approach for its proof.

It is not clear how useful Theorem 1.9 is for deciding whether an explicitly given expanding Thurston map is topologically conjugate to a rational map. It likely that our techniques can be used to formulate a more efficient criterion, but we will not pursue this further here. We content ourselves with a simple statement that easily follows from our results.

Theorem 1.10. *Let $f: S^2 \rightarrow S^2$ be a Thurston map with $\# \text{post}(f) = 3$. Then f is Thurston equivalent to a rational Thurston map. If the map f is expanding, then f is topologically conjugate to a rational Thurston map if and only if f has no periodic critical points.*

Theorem 1.2 can be used to study the topological and measure theoretic dynamics of an expanding Thurston map $f: S^2 \rightarrow S^2$ under iteration. For example, we have $h_{\text{top}}(f) = \log(\deg(f))$, where $h_{\text{top}}(f)$ is the topological entropy of f (Corollary 20.8). The following statement gives information of the statistical behavior of f under iteration.

Theorem 1.11. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map. Then there exists a unique measure μ of maximal entropy for f . The map f is mixing for μ .*

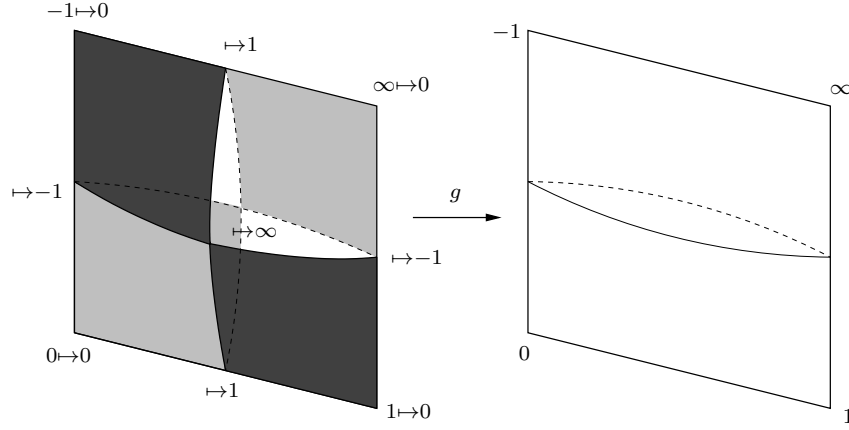
This theorem follows from results due to Haïssinsky-Pilgrim [HP09, Thm. 3.4.1]. We will present a different proof and give an explicit description of μ in terms of the cell decompositions $\mathcal{D}^n(F, \mathcal{C})$, where \mathcal{C} is an invariant curve as in Theorem 1.2 and $F = f^n$ (see Proposition 20.7 and Theorem 20.9).

If the map f has no periodic critical points, then the measure μ is *Ahlfors regular*. More precisely, if d is a visual metric for f with expansion factor $\Lambda > 1$, then for all balls with small radius r we have

$$\mu(B_d(x, r)) \asymp r^Q,$$

where $Q = \log(\deg(f))/\log(\Lambda)$ (Proposition 20.10). In particular, this number Q is the Hausdorff dimension of (S^2, d) .

1.2. Lattès maps. The simplest expanding Thurston maps are *Lattès maps*. They were the first examples of rational maps whose Julia set

FIGURE 1. The Lattès map g .

is the whole sphere. We will remind the reader of the construction by discussing a particular Lattès map in detail (see [La] and [Mi06]).

The unit square $[0, 1]^2 \subset \mathbb{R}^2 \cong \mathbb{C}$ can be conformally mapped to the upper half-plane in $\widehat{\mathbb{C}}$ such that the vertices $0, 1, 1 + i, i$ of the square correspond to the points $0, 1, \infty, -1$, respectively. By Schwarz reflection we can extend this to a map $\wp: \mathbb{C} \rightarrow \widehat{\mathbb{C}}$. Up to post-composition by a Möbius transformation this map is the classical *Weierstraß \wp -function*; it is doubly periodic with respect to the lattice $L := 2\mathbb{Z}^2$ and gives a double branched covering map of the torus $\mathbb{T}^2 := \mathbb{C}/L$ to the sphere $\widehat{\mathbb{C}}$.

Consider the map

$$\psi: \mathbb{C} \rightarrow \mathbb{C}, \quad u \mapsto \psi(u) := 2u.$$

One can check that there is a well-defined and unique map $g: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ such that the diagram

$$(1.2) \quad \begin{array}{ccc} \mathbb{C} & \xrightarrow{\psi} & \mathbb{C} \\ \wp \downarrow & & \downarrow \wp \\ \widehat{\mathbb{C}} & \xrightarrow{g} & \widehat{\mathbb{C}} \end{array}$$

commutes. The map g is rational, in fact

$$g(z) = 4 \frac{z(1 - z^2)}{(1 + z^2)^2} \quad \text{for } z \in \widehat{\mathbb{C}}.$$

The Julia set of g is the whole sphere.

One can describe g geometrically as follows. There is an essentially unique orbifold metric on $\widehat{\mathbb{C}}$ (see, for example, [Mi, App. E], and [McM,

App. A] for the terminology) with four conical singularities whose pull-back by \wp is the Euclidean metric on \mathbb{C} . Geometrically, the sphere equipped with this metric looks like a pillow. In general, a *pillow* is a metric space obtained from glueing two identical Euclidean polygons together along their boundary, equipped with the induced path-metric. In our case, the upper and lower half-planes in $\widehat{\mathbb{C}}$ equipped with the orbifold metric are isometric to copies of the square $[0, 1]^2$. If we glue two copies of this square along their boundaries, then we obtain the pillow. We color one of these squares, say the upper half-plane, white, and the other square, the lower half-plane, black. We divide each of these two squares in 4 smaller squares of half the side length, and color the 8 small squares in a checkerboard fashion black and white. If we map one such small white square to the large white square by a Euclidean similarity, then this map extends by reflection to the whole pillow. There are obviously many different ways to color and map the small squares. If we do this in an appropriate way as indicated in Figure 1, then we obtain the map g .

The vertices where four small squares intersect are the critical points of g . They are mapped by g to the set $\{1, \infty, -1\}$, which in turn is mapped to $\{0\}$. The point 0 is a fixed point of g . Hence g is a postcritically-finite map with $\text{post}(g) = \{0, 1, \infty, -1\}$. So the postcritical points of g are the vertices of the pillow, which are, in more technical terms, the conical singularities of our orbifold metric. For the map g we can take $n = 1$ in Theorem 1.2, and as the quasicircle \mathcal{C} the extended real line $\widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. Note that $\widehat{\mathbb{R}}$ contains the set $\{0, 1, \infty, -1\}$ of postcritical points of g . The set $g^{-1}(\widehat{\mathbb{R}})$ is the set of all edges of the small squares on the left hand side of Figure 1, and so the tiling in the picture is determined by $g^{-1}(\widehat{\mathbb{R}})$.

Theorem 1.2 enables us to give a combinatorial description as in this example for some iterate of *every* expanding Thurston map $f: S^2 \rightarrow S^2$. Indeed, suppose $F = f^n$ is an iterate of f for which there exists an invariant Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) = \text{post}(F) \subset \mathcal{C}$. Then we get a natural cell decomposition $\mathcal{D}^0 = \mathcal{D}^0(F, \mathcal{C})$ of S^2 , consisting of the points in $\text{post}(F)$ as vertices, the closures of the components of $\mathcal{C} \setminus \text{post}(F)$ as edges, and the closures of the two components of $S^2 \setminus \mathcal{C}$ as 2-dimensional cells or tiles. The cell-decomposition \mathcal{D}^0 pulls back under F (see Lemma 5.4) to a cell decomposition $\mathcal{D}^1 = \mathcal{D}^1(F, \mathcal{C})$ of S^2 such that F is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$, i.e., for each cell c in \mathcal{D}^1 the map $F|_c$ is a homeomorphism of c onto a cell in \mathcal{D}^0 . Moreover, since \mathcal{C} is F -invariant, \mathcal{D}^1 is a refinement of \mathcal{D}^0 . The pair $(\mathcal{D}^1, \mathcal{D}^0)$ together with a normalization essentially determines F uniquely (see Proposition 12.5

and Lemma 12.7). In this sense, F is described by finite combinatorial data.

1.3. Related work. Part of this paper has overlap with work by other researchers, notably Haïssinsky-Pilgrim [HP09], and Cannon-Floyd-Parry [CFP07]. Theorem 1.1 was announced by the first author during an Invited Address at the AMS Meeting at Athens, Ohio, in March 2004, where he gave a short outline of the proof. After the talk he was informed by W. Floyd that related results had independently obtained by Cannon-Floyd-Parry (which later appeared as [CFP07]).

Theorem 1.9 has already been published by Haïssinsky-Pilgrim as part of a more general statement [HP09, Thm. 4.2.11]. Special cases go back to work by the second author [Me02] and unpublished joint work by B. Kleiner and the first author. The current more general version seems to have emerged after a visit of the first author at the University of Indiana at Bloomington in February 2003.

During this visit the first author explained concepts of quasiconformal geometry to K. Pilgrim and his joint work with B. Kleiner on Cannon's conjecture in geometric group theory. K. Pilgrim in turn pointed out Theorem 10.4 and the ideas for its proof to the first author. After this visit versions of Theorem 1.9 with an outline for the proof were found independently by K. Pilgrim and the first author.

A proof of Theorem 1.9 was discovered soon afterwards by the authors using ideas from [Me02] (see [Me10] for an argument along similar lines) in combination with Theorem 1.2.

1.4. Outline of the paper and main ideas. The paper is organized as follows. After fixing some notation in Section 2, we review Thurston maps and some basic related concepts in Section 3. We also give a precise definition of an *expanding Thurston map* (Definition 3.2).

We then collect general facts about cell decompositions in Section 4. In particular, we introduce the concept of a *cellular Markov partition* (Definition 4.8). We will later show in Section 9 that under some additional assumptions a continuous map on a compact metric space is a factor of a subshift of finite type if it has a cellular Markov partition (Proposition 9.1).

In Section 5 we specialize to cell decompositions on 2-spheres. Lemma 5.2 shows how to construct branched covering maps and Thurston maps from cell decompositions. One can pull-back a cell decomposition \mathcal{D} of a 2-sphere by a branched covering map f if the vertex set \mathbf{V} of \mathcal{D} contains the critical values of f (Lemma 5.4). Since the set $\text{post}(f)$ contains the critical values of all iterates of a Thurston map f , this applies to all iterates f^n if $\text{post}(f) \subset \mathbf{V}$.

We use this in Section 6 to show that if $f: S^2 \rightarrow S^2$ is a Thurston map and $\mathcal{C} \subset S^2$ a Jordan curve with $\text{post}(f) \subset \mathcal{C}$, then we obtain a natural sequence $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ of cell decompositions of S^2 . These cell decompositions are our most important technical tool for studying Thurston maps. Their properties are summarized in Proposition 6.1. Simple applications are Proposition 6.3, which gives a classification of all Thurston maps f with $\#\text{post}(f) = 2$ up to Thurston equivalence, and Corollary 6.4 showing that for an expanding Thurston map f we have $\#\text{post}(f) \geq 3$.

In general, the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ are not compatible for different levels n unless the Jordan curve \mathcal{C} is f -invariant. To overcome the ensuing problems, we introduce the concept of an n -flower $W^n(p)$ of a vertex p in the cell decomposition \mathcal{D}^n (Section 7). The set $W^n(p)$ is formed by the interiors of all cells in \mathcal{D}^n that meet p (see Definition 7.1 and Lemma 7.2). An important fact is that while in general a component of the preimage $f^{-n}(K)$ of a small connected set K will not be contained in an n -tile (i.e., a 2-dimensional cell in \mathcal{D}^n), it is always contained in an n -flower (Lemma 7.8).

In Section 7 we also define a quantity $D_n = D_n(f, \mathcal{C})$ that measures the combinatorial expansion rate of a Thurston map. It is given by the minimal number of n -tiles needed to form a connected set joining “opposite sides” of \mathcal{C} (see (7.4) and Definition 7.6).

Visual metrics for expanding Thurston maps are introduced in Section 8. Their most important properties are stated in Proposition 8.9 and Lemma 8.10. In particular, if $f: S^2 \rightarrow S^2$ is an expanding Thurston map, and d a visual metric, then the d -diameters of cells in \mathcal{D}^n will approach 0 at an exponential rate as $n \rightarrow \infty$. This implies that lifts of paths under f^n shrink to 0 exponentially fast if $n \rightarrow \infty$ (Lemma 8.11). This fact is of crucial importance. It implies that every expanding Thurston map is a factor of a shift operator (see Section 9 where Theorem 1.6 is proved).

The exponential shrinking of lifts will also be used in Section 10 to show that if two expanding Thurston maps are Thurston equivalent, then they are topologically conjugate (Theorem 10.4). The idea for the proof is to lift an initial isotopy repeatedly to obtain a sequence of isotopies that form a “tower”. If the Thurston maps are expanding, then the diameters of the tracks of these isotopies shrink fast enough so that the isotopies converge to a time independent homeomorphism that provides the desired conjugacy. In this section we also prove some results on isotopies of Jordan curves.

In the next Section 11 we study the cell decomposition $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ under the additional assumption that \mathcal{C} is f -invariant. In this case

\mathcal{D}^{n+k} is a refinement of \mathcal{D}^n for all $n, k \in \mathbb{N}_0$; on a more intuitive level, each cell in any of the cell decompositions \mathcal{D}^n is “subdivided” by cells on a higher level. Moreover, the pair $(\mathcal{D}^{n+k}, \mathcal{D}^n)$ is a cellular Markov partition for f (Proposition 11.1). If a Thurston map has an invariant Jordan curve \mathcal{C} with $\text{post}(f) \subset \mathcal{C}$, then it can be described by a *two-tile subdivision rule* (Definition 12.4) as discussed in Section 12. The main result here is Proposition 12.5 that gives a general method for constructing Thurston maps with invariant curves from a two-tile subdivision rule (see Definition 12.1 and the following discussion).

We call f *combinatorially expanding* for an invariant curve \mathcal{C} if there exists a number $n_0 \in \mathbb{N}$ such that $D_{n_0}(f, \mathcal{C}) \geq 2$ (Definition 11.4). This means that there exists $n_0 \in \mathbb{N}$ so that no n_0 -tile joins opposite sides of \mathcal{C} . In this case the numbers $D_n = D_n(f, \mathcal{C})$ grow at an exponential rate as $n \rightarrow \infty$ (Lemma 11.3). It is easy to see that if f is an expanding Thurston map, then $D_n(f, \mathcal{C}) \rightarrow \infty$ as $n \rightarrow \infty$. In particular, if \mathcal{C} is f -invariant, then f is combinatorially expanding for \mathcal{C} . The converse is not true in general, but in Section 13 we show that every combinatorially expanding Thurston map is (Thurston) equivalent to an expanding Thurston map (Proposition 13.1 and Corollary 13.18).

The intuitive reason for this is that if the diameters of the tiles in $\mathcal{D}^n(f, \mathcal{C})$ fail to shrink to zero as $n \rightarrow \infty$, then one can “correct” the map f so that this becomes true without affecting the combinatorics of the cell decompositions $\mathcal{D}^n(f, \mathcal{C})$.

It is somewhat cumbersome to implement this idea. We do this by introducing an equivalence relation that forces descending sequences of n -tiles to shrink to points as $n \rightarrow \infty$. We then invoke Moore’s theorem (Theorem 13.4) to show that the quotient space of the original 2-sphere S^2 by this relation is also a 2-sphere \tilde{S}^2 . The original Thurston map $f: S^2 \rightarrow S^2$ descends to a map $\tilde{f}: \tilde{S}^2 \rightarrow \tilde{S}^2$ and one can show that \tilde{f} is an expanding Thurston map that is equivalent to the original map f .

Section 14 contains some auxiliary statements on graphs. The main result is Lemma 14.5 that gives a sufficient criterion when a Jordan curve can be isotoped into the 1-skeleton of a cell decomposition of a 2-sphere.

Existence and uniqueness results for invariant Jordan curves are proved in Section 15. This section constitutes the center of the present work. Here we establish Theorems 1.2, 1.3, and 1.4, and Corollary 1.5. Among these statements Theorem 1.2 is the most difficult to prove. It is based on Lemma 14.5 and Theorem 1.3.

Roughly speaking, the idea for the proof of Theorem 1.2 can be summarized as follows. Let $f: S^2 \rightarrow S^2$ be an expanding Thurston

map, and $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$. Consider the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ of S^2 induced by f and \mathcal{C} . If n is large enough, one can find a homeomorphism ψ of S^2 that maps \mathcal{C} into the 1-skeleton $f^{-n}(\mathcal{C})$ of \mathcal{D}^n and is isotopic to the identity on S^2 by an isotopy that fixes the points in $\text{post}(f)$. Essentially, this follows from the fact that f is expanding and so the 1-skeleton $f^{-n}(\mathcal{C})$ of \mathcal{D}^n forms a fine “grid” in S^2 for n large. This grid contains $\text{post}(f)$ and allows us to trace the curve \mathcal{C} closely.

The map $\widehat{F} = \psi \circ f^n$ will be a Thurston map with an invariant curve $\widehat{\mathcal{C}} = \psi(\mathcal{C}) \supseteq \text{post}(\widehat{F}) = \text{post}(f)$. Moreover, if n is large enough, then \widehat{F} is combinatorially expanding. Actually, we can even assume that \widehat{F} is expanding, because if necessary, the map can be “corrected” to have this property (Corollary 13.18). The expanding Thurston maps $F = f^n$ and \widehat{F} are Thurston equivalent and hence topologically conjugate by Theorem 10.4. Since the map \widehat{F} has the invariant Jordan curve $\widehat{\mathcal{C}}$, one obtains an F -invariant curve with the desired properties as an image of $\widehat{\mathcal{C}}$ under a homeomorphism that conjugates \widehat{F} and F .

In Section 16 we review the notion of a quasicircle and prove Theorem 1.8. We also show that if $f: S^2 \rightarrow S^2$ is an expanding Thurston map, S^2 is equipped with a visual metric d for S^2 , and the cell decompositions $\mathcal{D}^n(f, \mathcal{C})$, $n \in \mathbb{N}_0$, are obtained from an f -invariant Jordan curve \mathcal{C} , then the edges in these cell decompositions are uniform quasiarcs and the boundaries of tiles are uniform quasicircles (Proposition 16.2).

If $f: S^2 \rightarrow S^2$ is an expanding Thurston map and S^2 is equipped with a visual metric d for f , then (S^2, d) has some properties that are important in the analysis of metric spaces. For example, (S^2, d) is linearly locally connected (Proposition 16.3). Moreover, (S^2, d) is a doubling metric space if and only if $f: S^2 \rightarrow S^2$ has no periodic critical points (Theorem 17.2). Actually, the absence of periodic critical points even implies that (S^2, d) is Ahlfors regular (Proposition 20.10).

In Section 18 we revisit visual metrics. We introduce the combinatorial expansion factor $\Lambda_0(f)$ associated with an expanding Thurston map f and prove Theorem 1.7.

The topic of Section 19 are rational Thurston maps on the Riemann sphere $\widehat{\mathbb{C}}$. For such maps the notion of expansion can be characterized in more familiar terms (Proposition 19.1). Here we prove Theorems 1.1, 1.9, and 1.10.

In Section 20 we study the dynamics of an expanding Thurston map from a measure theoretic point of view. The main result is Theorem 20.9, which gives an explicit description of the unique measure of maximal entropy.

We list some open problems in the final Section 21.

Acknowledgements. The authors would like to thank J. Cannon, W. Floyd, P. Haïssinsky, B. Kleiner, W. Parry, K. Pilgrim, J. Souto, D. Sullivan, Q. Yin, and M. Zieve for interesting discussions and useful comments.

1.5. Examples of Thurston maps. Throughout the paper we consider several examples of Thurston maps to illustrate various phenomena. We list them here with a short description for easy reference. The relevant terms used in these descriptions are later defined in the body of the paper.

A Lattès map was considered in Section 1.2.

The map in Example 12.10 is $f_1(z) = z^2 - 1$; it realizes a two-tile subdivision rule that is not combinatorially expanding.

In Example 12.11 there are two maps f_2 and \tilde{f}_2 that both realize the barycentric subdivision rule. The map f_2 is a rational map, but it is not expanding (i.e., its Julia set is not the whole Riemann sphere $\hat{\mathbb{C}}$). The map \tilde{f}_2 however is expanding. It is an example of an expanding Thurston map with periodic critical points.

The map f_3 in Example 12.12 (realizing a certain two-tile subdivision rule) is an obstructed map. This means f_3 is not Thurston equivalent to a rational map.

The map f_4 in Example 12.13 is again not Thurston equivalent to a rational map. While somewhat easier than the map f_3 in Example 12.12, it is less generic, since f_4 has a parabolic orbifold, whereas f_3 has a hyperbolic orbifold. The map f_4 realizes the 2-by-3 subdivision rule. With respect to a suitable visual metric for f_4 , the sphere S^2 consists of two copies of a Rickman's rug.

In Example 12.14 a whole class of maps is considered. The first one is the map $f_5(z) = 1 - 2/z^2$ which realizes a simple two-tile subdivision rule. By “adding flaps” we obtain the other maps. All these maps are rational; in fact they are given by an explicit formula, which makes them easy to understand and visualize.

In Example 13.20 we consider a Thurston map f that is not combinatorially expanding, yet Thurston equivalent to an expanding Thurston map g . This shows that the sufficient condition in Proposition 13.1 is not necessary.

In Example 15.1 we illustrate the main ideas of Section 15. In particular, we show how for a specific map f an f -invariant curve $\tilde{\mathcal{C}}$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$ is constructed; see Figure 18.

In Example 15.4 we show that the Lattès map g from Section 1.2 has infinitely many distinct g -invariant curves \mathcal{C} with $\text{post}(g) \subset \mathcal{C}$.

In Example 15.5 we consider an expanding Thurston map f for which no f -invariant Jordan curve \mathcal{C} with $\text{post}(f) \subset \mathcal{C}$ exists.

In Remark 15.10 we show that the f^n -invariant curve $\tilde{\mathcal{C}}$ given by Theorem 1.2 will in general depend on n .

In Example 15.11 we use another Lattès map to illustrate an iterative construction of invariant curves; see Figure 21.

Example 15.17 shows what can happen if one of the necessary conditions in the iterative procedure for producing invariant curve is violated. Namely, the “limiting object” $\tilde{\mathcal{C}}$ is not a Jordan curve anymore. The map used to illustrate this phenomenon is again a Lattès map.

In Example 15.18 (again a Lattès map) we obtain a non-trivial (in particular non-smooth) invariant curve that is rectifiable.

In Example 18.5 we exhibit an expanding Thurston map f for which no visual metric with expansion factor Λ equal to the combinatorial expansion factor $\Lambda_0(f)$ exists. This shows that part (ii) in Theorem 1.7 cannot be improved.

2. NOTATION

We denote by $\mathbb{N} = \{1, 2, \dots\}$ the set of natural numbers, and by $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ the set of natural numbers including 0. The symbol i stands for the imaginary unit in the complex plane \mathbb{C} . We define $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$ as the open unit disk in \mathbb{C} .

Let (X, d) be a metric space, $a \in X$ and $r > 0$. We denote by $B_d(a, r) = \{x \in X : d(a, x) < r\}$ and by $\overline{B}_d(a, r) = \{x \in X : d(a, x) \leq r\}$ the open and the closed ball of radius r centered at a , respectively. If $A, B \subset X$, we let $\text{diam}_d(A)$ be the diameter, \overline{A} be the closure of A in X , and

$$\text{dist}_d(A, B) = \inf\{d(x, y) : a \in A, y \in B\}$$

be the distance of A and B . If $p \in X$, we let $\text{dist}_d(p, A) = \text{dist}_d(\{p\}, A)$. If $\epsilon > 0$ then

$$\mathcal{N}_d^\epsilon(A) := \{x \in X : \text{dist}_d(x, A) < \epsilon\}$$

is the open ϵ -neighborhood of A . We drop the subscript d in $B_d(a, r)$, etc., if the metric d is clear from the context.

The cardinality of a set M is denoted by $\#M$ and the identity map on M by id_M . If $f : M \rightarrow M$ is a map, then f^n for $n \in \mathbb{N}$ is the n -th iterate of f . We set $f^0 := \text{id}_M$.

Two non-negative quantities a and b are said to be *comparable* if there is a constant $C \geq 1$ depending on some obvious ambient parameters

such that

$$\frac{1}{C}a \leq b \leq Ca.$$

We then write $a \asymp b$. The constant C is referred to by $C(\asymp)$. We write $a \lesssim b$ or $b \gtrsim a$, if there is a constant $C > 0$ such that $a \leq Cb$. We refer to the constant as $C(\lesssim)$ or $C(\gtrsim)$.

3. THURSTON MAPS

Let S^2 be a topological 2-sphere with a fixed orientation, and $f: S^2 \rightarrow S^2$ be a continuous map. Then f is called a *branched covering map* of S^2 if we can write it locally as the map $z \mapsto z^d$ for some $d \in \mathbb{N}$ after orientation-preserving homeomorphic changes of coordinates in domain and range. More precisely, we require that for each point $p \in S^2$ there exists $d \in \mathbb{N}$, open neighborhoods U of p and V of $q = f(p)$, open neighborhoods U' and V' of $0 \in \widehat{\mathbb{C}}$ and orientation-preserving homeomorphisms $\varphi: U \rightarrow U'$ and $\psi: V \rightarrow V'$ with $\varphi(p) = 0$ and $\psi(q) = 0$ such that

$$(\psi \circ f \circ \varphi^{-1})(z) = z^d$$

for all $z \in U'$.

The integer $d =: \deg_f(p) \geq 1$ is uniquely determined by f and p , and called the *local degree* of the map f at p . A point $c \in S^2$ with $\deg_f(c) \geq 2$ is called a *critical point* of f , and a point that has a critical point as a preimage a *critical value*. The set of all critical points is denoted by $\text{crit}(f)$. Obviously, if f is a branched covering map on S^2 , then $\text{crit}(f)$ only consists of isolated points and is hence a finite subset of S^2 . Moreover, f is an open and surjective mapping, and *finite-to-one*, i.e., every point has finitely many preimages under f . More precisely, if $\deg(f)$ is the topological degree of f , then

$$\sum_{p \in f^{-1}(q)} \deg_f(p) = \deg(f)$$

for every $q \in S^2$ (see [Ha, Sect. 2.2]). In particular, if q is not a critical value of f , then q has precisely $\deg(f)$ preimages.

For $n \in \mathbb{N}_0$ we denote by f^n the n -th iterate of f (where $f^0 := \text{id}_{S^2}$). If f is a branched covering map on S^2 , then the same is true for f^n and we have $\deg(f^n) = \deg(f)^n$ and

$$(3.1) \quad \text{crit}(f^n) = \text{crit}(f) \cup f^{-1}(\text{crit}(f)) \cup \dots \cup f^{-(n-1)}(\text{crit}(f)).$$

The *set of postcritical points* of f is defined as

$$\text{post}(f) := \bigcup_{n \in \mathbb{N}} \{f^n(c) : c \in \text{crit}(f)\}.$$

If the cardinality $\#\text{post}(f)$ is finite, then f is called *postcritically-finite*.

For $n \in \mathbb{N}$ we have $\text{post}(f^n) = \text{post}(f)$ and $f^n(\text{crit}(f^n)) \subset \text{post}(f)$. The last inclusion implies that away from $\text{post}(f)$ each iterate f^n is a covering map and all “branches of the inverse of f^n ” are defined; more precisely, if $U \subset S^2 \setminus \text{post}(f)$ is a path connected and simply connected set, $q \in U$, and $p \in S^2$ a point with $f^n(p) = q$, then there exists a unique continuous map $g: U \rightarrow S^2$ with $g(q) = p$ and $f^n \circ g = \text{id}_U$. Informally, we refer to such a right inverse of f^n as a “branch of f^{-n} ”.

We can now record the definition of the main object of investigation in this paper.

Definition 3.1 (Thurston maps). A *Thurston map* is a branched covering map $f: S^2 \rightarrow S^2$ of a 2-sphere S^2 with $\deg(f) \geq 2$ and finite set of postcritical points.

There are no Thurston maps with $\#\text{post}(f) \in \{0, 1\}$ (see Remark 5.5), and all Thurston maps with $\#\text{post}(f) = 2$ are Thurston equivalent (see below) to a map $z \mapsto z^k$, $k \in \mathbb{Z} \setminus \{-1, 0, 1\}$, on the Riemann sphere (see Proposition 6.3).

Let $f: S^2 \rightarrow S^2$ be a Thurston map and \mathcal{C} be a Jordan curve in S^2 with $\text{post}(f) \subset \mathcal{C}$. We fix a metric d on S^2 that induces the standard topology on S^2 . For $n \in \mathbb{N}$ we denote by $\text{mesh}(f, n, \mathcal{C})$ the supremum of the diameters of all connected components of the set $f^{-n}(S^2 \setminus \mathcal{C})$.

Definition 3.2 (Expansion). A Thurston map $f: S^2 \rightarrow S^2$ is called *expanding* if there exists a Jordan curve \mathcal{C} in S^2 with $\text{post}(f) \subset \mathcal{C}$ and

$$(3.2) \quad \lim_{n \rightarrow \infty} \text{mesh}(f, n, \mathcal{C}) = 0.$$

As we will see later, the set $f^{-n}(S^2 \setminus \mathcal{C})$ has only finitely many components, so the supremum in the definition of $\text{mesh}(f, n, \mathcal{C})$ is actually a maximum.

We will also prove (see Lemma 8.1) that if the relation (3.2) is satisfied for one Jordan curve $\mathcal{C} \supset \text{post}(f)$, then it actually holds for every such curve.

Note further that this really is a topological property, as it is independent of the choice of the metric on S^2 if it induces the given topology on S^2 . Our notion of expansion for a Thurston map is equivalent to a similar concept of expansion introduced by Haïssinsky-Pilgrim (see [HP09, Sect. 2.2] and Proposition 8.2).

If the Thurston map is a rational map $f: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ on the Riemann sphere $\widehat{\mathbb{C}}$, then one can show that f is expanding if and only if f does

not have periodic critical points if and only if its Julia set is equal to $\widehat{\mathbb{C}}$ (see Section 19 for more details).

To define a suitable notion of equivalence for Thurston maps we recall the definition of an isotopy between spaces. Let $I = [0, 1]$, and X and Y be topological spaces. An *isotopy between X and Y* is a continuous map $\phi: X \times I \rightarrow Y$ such that each map $\phi_t := \phi(\cdot, t)$, $t \in I$, is a homeomorphism of X onto Y .

For a subset A of X , we say ϕ is an isotopy *relative to A* (abbreviated “ ϕ is an isotopy rel. A ”) if $\phi_t(a) = \phi_0(a)$ for all $a \in A$ and $t \in I$. So this means that the image of each point in A remains fixed during the isotopy. Two homeomorphisms $\varphi, \psi: X \rightarrow Y$ are called *isotopic rel. A* if there exists an isotopy $\phi: X \times I \rightarrow Y$ rel. A with $\phi_0 = \varphi$ and $\phi_1 = \psi$.

Definition 3.3 (Thurston equivalence). Two Thurston maps $f: S^2 \rightarrow S^2$ and $g: \widehat{S}^2 \rightarrow \widehat{S}^2$ are called (*Thurston*) *equivalent* if there exist homeomorphisms $h_0, h_1: S^2 \rightarrow \widehat{S}^2$ that are isotopic rel. $\text{post}(f)$ and satisfy $h_0 \circ f = g \circ h_1$.

Here \widehat{S}^2 is another 2-sphere. Often $S^2 = \widehat{S}^2$, but sometimes it is important to distinguish the spheres on which the Thurston maps are defined.

For equivalent Thurston maps as in Definition 3.3 we have the following commutative diagram:

$$(3.3) \quad \begin{array}{ccc} S^2 & \xrightarrow{h_1} & \widehat{S}^2 \\ f \downarrow & & \downarrow g \\ S^2 & \xrightarrow{h_0} & \widehat{S}^2. \end{array}$$

Note that in this situation

$$(3.4) \quad \text{post}(g) = h_0(\text{post}(f)) = h_1(\text{post}(f)).$$

Indeed, it is clear that

$$(3.5) \quad \text{crit}(g) = h_1(\text{crit}(f)).$$

Moreover, since $h_0|_{\text{post}(f)} = h_1|_{\text{post}(f)}$ and $f^n(\text{crit}(f)) \subset \text{post}(f)$ for all $n \in \mathbb{N}$, the relation (3.5) inductively implies

$$g^n(\text{crit}(g)) = h_0(f^n(\text{crit}(f))) = h_1(f^n(\text{crit}(f)))$$

for all $n \in \mathbb{N}$. Hence

$$\begin{aligned} \text{post}(g) &= \bigcup_{n \in \mathbb{N}} g^n(\text{crit}(g)) = \bigcup_{n \in \mathbb{N}} h_0(f^n(\text{crit}(f))) \\ &= h_0(\text{post}(f)) = h_1(\text{post}(f)) \end{aligned}$$

as desired.

We call the maps f and g *topologically conjugate* if there exists a homeomorphism $h: S^2 \rightarrow \widehat{S}^2$ such that $h \circ f = g \circ h$. We will see later (see Theorem 10.4) that if two expanding Thurston maps are equivalent, then they are topologically conjugate.

If S^2 and \widehat{S}^2 are 2-spheres, $f: S^2 \rightarrow S^2$ is an expanding Thurston map and $g: \widehat{S}^2 \rightarrow \widehat{S}^2$ is a topologically conjugate to f , then g is also an expanding Thurston map. So the notion of an expanding Thurston map is invariant under topological conjugacy.

4. CELL DECOMPOSITIONS

In this section we review some facts about cell decompositions. Most of this material is fairly standard. For the purpose of the present paper we could have restricted ourselves to cell decompositions of subsets of a 2-sphere, but it is more transparent to discuss the topic in greater generality. At first reading the reader may want to skim through this section to pick up relevant definitions and statements.

A crucial concept introduced in this section is the notion of a “cellular Markov partition” (Definition 4.8) of a map. One can show that if a map admits a cellular Markov partition, then its dynamics is linked to symbolic dynamics; in particular, it is a factor of a subshift of finite type (see Proposition 9.1).

In the following X will always be a locally compact Hausdorff space. A (*closed topological*) *cell of dimension* $n = \dim(c) \in \mathbb{N}$ in X is a set $c \subset X$ that is homeomorphic to the closed unit ball $\overline{\mathbb{B}}^n$ in \mathbb{R}^n . We denote by ∂c the set of points corresponding to $\partial \overline{\mathbb{B}}^n$ under such a homeomorphism between c and $\overline{\mathbb{B}}^n$. This is independent of the homeomorphism chosen, and the set ∂c is well-defined. We call ∂c the *boundary* and $\text{int}(c) = c \setminus \partial c$ the *interior* of c . Note that boundary and interior of c in this sense will in general not agree with the boundary and interior of c regarded as a subset of the topological space X . A *cell of dimension 0* in X is a set $c \subset X$ consisting of a single point. We set $\partial c = \emptyset$ and $\text{int}(c) = c$ in this case.

Definition 4.1 (Cell decompositions). Suppose that \mathcal{D} is a collection of cells in a locally compact Hausdorff space X . We say that \mathcal{D} is a *cell decomposition* of X provided the following conditions are satisfied:

- (i) the union of all cells in \mathcal{D} is equal to X ,
- (ii) we have $\text{int}(\sigma) \cap \text{int}(\tau) = \emptyset$, whenever $\sigma, \tau \in \mathcal{D}$, $\sigma \neq \tau$,
- (iii) if $\tau \in \mathcal{D}$, then $\partial \tau$ is a union of cells in \mathcal{D} ,

- (iv) every point in X has a neighborhood that meets only finitely many cells in \mathcal{D} .

If \mathcal{D} is a collection of cells in some ambient space X , then we call \mathcal{D} a *cell complex* if \mathcal{D} is a cell decomposition of the underlying set

$$|\mathcal{D}| = \bigcup \{c : c \in \mathcal{D}\}.$$

Suppose \mathcal{D} is a cell decomposition of X . By (iv), every compact subset of X can only meet finitely many cells in \mathcal{D} . In particular, if X is compact, then \mathcal{D} consists of only finitely many cells. Moreover, for each $\tau \in \mathcal{D}$, the set $\partial\tau$ is compact and hence equal to a finite union of cells in \mathcal{D} . It follows from basic dimension theory that if $\dim(\tau) = n$, then $\partial\tau$ is equal to a union of cells in \mathcal{D} that have dimension $n - 1$.

The union X^n of all cells in \mathcal{D} of dimension $\leq n$ is called the *n-skeleton* of the cell decomposition. It is useful to set $X^{-1} = \emptyset$. It follows from the local compactness of X and property (iv) of a cell decomposition that X^n is a closed subset of X for each $n \in \mathbb{N}_0$.

By the last remark in the previous paragraph, we have $\partial\tau \subset X^{n-1}$ for each $\tau \in \mathcal{D}$ with $\dim(\tau) = n$.

Lemma 4.2. *Let \mathcal{D} be a cell decomposition of X . Then for each $n \in \mathbb{N}_0$ the n -skeleton X^n is equal to the disjoint union of the sets $\text{int}(c)$, $c \in \mathcal{D}$, $\dim(c) \leq n$. Moreover, X is equal to the disjoint union of the sets $\text{int}(c)$, $c \in \mathcal{D}$.*

Proof. We show the first statement by induction on $n \in \mathbb{N}_0$. Since $\text{int}(c) = c$ for each cell c in \mathcal{D} of dimension 0, it is clear that X^0 is the disjoint union of the interiors of all cells $c \in \mathcal{D}$ with $\dim(c) = 0$.

Suppose that the statement is true for X^n , and let $p \in X^{n+1}$ be arbitrary. If $p \in X^n$, then p is contained in the interior of a cell $c \in \mathcal{D}$ with $\dim(c) \leq n$ by induction hypothesis. In the other case, $p \in X^{n+1} \setminus X^n$, and so there exists $c \in \mathcal{D}$ with $\dim(c) = n + 1$ and $p \in c$. Since $\partial c \subset X^n$, it follows that $p \in c \setminus \partial c = \text{int}(c)$. So X^{n+1} is the union of the interiors of all cells c in \mathcal{D} with $\dim(c) \leq n + 1$. This union is disjoint, because distinct cells in a cell decomposition have disjoint interiors.

The second statement follows from the first, and the obvious fact that $X = \bigcup_{n \in \mathbb{N}_0} X^n$. \square

The lemma immediately implies that if $\tau \in \mathcal{D}$ and $\dim(\tau) = n$, then each point $p \in \text{int}(\tau)$ is an interior point of τ regarded as a subset of the topological space X^n .

Lemma 4.3. *Let \mathcal{D} be a cell decomposition of X .*

- (i) If σ and τ are two distinct cells in \mathcal{D} with $\sigma \cap \tau \neq \emptyset$, then one of the following statements holds: $\sigma \subset \partial\tau$, $\tau \subset \partial\sigma$, or $\sigma \cap \tau = \partial\sigma \cap \partial\tau$ and this intersection consists of cells in \mathcal{D} of dimension strictly less than $\min\{\dim(\sigma), \dim(\tau)\}$.
- (ii) If $\sigma, \tau_1, \dots, \tau_n$ are cells in \mathcal{D} and $\text{int}(\sigma) \cap (\tau_1 \cup \dots \cup \tau_n) \neq \emptyset$, then $\sigma \subset \tau_i$ for some $i \in \{1, \dots, n\}$.

Proof. (i) We may assume that $l = \dim(\sigma) \leq m = \dim(\tau)$, and prove the statement by induction on m . The case $m = 0$ is vacuous and hence trivial. Assume that the statement is true whenever both cells have dimension $< m$. If $l = m$ then by definition of a cell decomposition $\text{int}(\sigma)$ is disjoint from $\tau \subset \text{int}(\tau) \cup X^{m-1}$, and similarly $\text{int}(\tau) \cap \sigma = \emptyset$. Hence $\sigma \cap \tau = \partial\sigma \cap \partial\tau$. Moreover, both sets $\partial\sigma$ and $\partial\tau$ consist of finitely many cells in \mathcal{D} of dimension $\leq m-1$. Applying the induction hypothesis to pairs of these cells, we see that $\partial\tau \cap \partial\sigma$ consists of cells of dimension $< m$ as desired.

If $l < m$, then $\sigma \subset X^{m-1}$ and so $\sigma \cap \text{int}(\tau) = \emptyset$. This shows that $\sigma \cap \tau = \sigma \cap \partial\tau$. Moreover, we have $\partial\tau = c_1 \cup \dots \cup c_s$, where c_1, \dots, c_s are cells of dimension $m-1$. So we can apply the induction hypothesis to the pairs (σ, c_i) . If $\sigma = c_i$ or $\sigma \subset \partial c_i$ for some i , then $\sigma \subset \partial\tau$; we cannot have $c_i \subset \partial\sigma$, because c_i has dimension $m-1$, and $\partial\sigma$ is a set of topological dimension $< m-1$. So if none of the first possibilities occurs, then $\sigma \cap c_i = \emptyset$ or $\sigma \cap c_i = \partial\sigma \cap \partial c_i$ and this set consists of cells of dimension $< l$ (by induction hypothesis) contained in $\partial c_i \subset c_i \subset \partial\tau$ for all i . In this case $\sigma \cap \tau = \partial\sigma \cap \partial\tau$, and this sets consists of cells of dimension $< l$ as desired. The claim follows.

(ii) There exists $i \in \{1, \dots, n\}$ with $\text{int}(\sigma) \cap \tau_i \neq \emptyset$. By the alternatives in (i) we then must have $\sigma = \tau_i$ or $\sigma \subset \partial\tau_i$. Hence $\sigma \subset \tau_i$. \square

Lemma 4.4. *Let $A \subset X$ be a closed set, and $U \subset X \setminus A$ a nonempty open and connected set. If $\partial U \subset A$, then U is a connected component of $X \setminus A$.*

Proof. Since U is a nonempty connected set in the complement of A , this set is contained in a unique connected component V of $X \setminus A$. Since $\partial U \subset A \subset X \setminus V$, we have $V \cap \overline{U} = V \cap U = U$ showing that U is relatively open and closed in V . Since $U \neq \emptyset$ and V is connected, it follows that $U = V$ as desired. \square

Lemma 4.5. *Let \mathcal{D} be a cell decomposition of X with n -skeleton X^n , $n \in \{-1\} \cup \mathbb{N}_0$. Then for each $n \in \mathbb{N}_0$ the nonempty connected components of $X^n \setminus X^{n-1}$ are precisely the sets $\text{int}(\tau)$, $\tau \in \mathcal{D}$, $\dim(\tau) = n$.*

Proof. Let τ be a cell in \mathcal{D} with $\dim(\tau) = n$. Then $\text{int}(\tau)$ is a connected set contained in $X^n \setminus X^{n-1}$ that is relatively open with respect to X^n .

Its relative boundary is a subset of $\partial\tau$ and hence contained in the closed set X^{n-1} . It follows by Lemma 4.4 that $\text{int}(\tau)$ is equal to a component V of $X^n \setminus X^{n-1}$.

Conversely, suppose that V is a nonempty connected component of $X^n \setminus X^{n-1}$. Pick a point $p \in V$. Then p lies in the interior of a unique cell $\tau \in \mathcal{D}$ with $\dim(\tau) = n$. It follows from the first part of the proof that $V = \text{int}(\tau)$. \square

Definition 4.6 (Refinements). Let \mathcal{D}' and \mathcal{D} be two cell decomposition of the space X . We say that \mathcal{D}' is a *refinement* of \mathcal{D} if the following two conditions are satisfied:

- (i) For every cell $\sigma \in \mathcal{D}'$ there exists a cell $\tau \in \mathcal{D}$ with $\sigma \subset \tau$.
- (ii) Every cell $\tau \in \mathcal{D}$ is the union of all cells $\sigma \in \mathcal{D}'$ with $\sigma \subset \tau$.

It is easy to see that if \mathcal{D}' is a refinement of \mathcal{D} and $\tau \in \mathcal{D}$, then the cells $\sigma \in \mathcal{D}'$ with $\sigma \subset \tau$ form a cell decomposition of τ . Moreover, every cell in \mathcal{D}' arises in this way. So roughly speaking, the refinement \mathcal{D}' of the cell decomposition \mathcal{D} is obtained by decomposing each cell in \mathcal{D} into smaller cells. We informally refer to this process as *subdividing* the cells in \mathcal{D} by the smaller cells in \mathcal{D}' .

Lemma 4.7. *Let \mathcal{D}' and \mathcal{D} be two cell decompositions of X , and \mathcal{D}' be a refinement of \mathcal{D} . Then for every cell $\sigma \in \mathcal{D}'$ there exists a minimal cell $\tau \in \mathcal{D}$ with $\sigma \subset \tau$, i.e., if $\tilde{\tau} \in \mathcal{D}$ is another cell with $\sigma \subset \tilde{\tau}$, then $\tau \subset \tilde{\tau}$. Moreover, τ is the unique cell with $\text{int}(\sigma) \subset \text{int}(\tau)$.*

Proof. First note that if $\sigma \in \mathcal{D}'$, $\tau_1, \dots, \tau_n \in \mathcal{D}$ and

$$\text{int}(\sigma) \cap (\tau_1 \cup \dots \cup \tau_n) \neq \emptyset,$$

then $\sigma \subset \tau_i$ for some $i \in \{1, \dots, n\}$. Indeed, by definition of a refinement the union of all cells in \mathcal{D}' contained in some τ_i covers $\tau_1 \cup \dots \cup \tau_n$. Hence this union meets $\text{int}(\sigma)$. It follows from Lemma 4.3 (ii) that σ is contained in one of these cells from \mathcal{D}' and hence in one of the cells τ_i .

Now if $\sigma \in \mathcal{D}'$ is arbitrary, then σ is contained in some cell of \mathcal{D} by definition of a refinement, and hence in a cell $\tau \in \mathcal{D}$ of minimal dimension. Then τ is minimal among all cells in \mathcal{D} containing σ . Indeed, let $\tilde{\tau} \neq \tau$ be another cell in \mathcal{D} containing σ . We want to show that $\tau \subset \tilde{\tau}$.

One of the alternatives in Lemma 4.3 (i) occurs. If $\tau \subset \partial\tilde{\tau} \subset \tilde{\tau}$ we are done. The second alternative, $\tilde{\tau} \subset \partial\tau$, is impossible, since τ has minimal dimension among all cells containing σ . The third alternative leads to $\sigma \subset \tau \cap \tilde{\tau} = \partial\tau \cap \partial\tilde{\tau}$, where the latter intersection consists of cells in \mathcal{D} of dimension $< \dim(\tau)$. By the first part of

the proof σ is contained in one of these cells, again contradicting the definition of τ . Hence τ is minimal.

We have $\text{int}(\sigma) \subset \text{int}(\tau)$; for otherwise $\text{int}(\sigma)$ meets $\partial\tau$ which is a union of cells in \mathcal{D} . Then σ would be contained in one of these cells by the first part of the proof. This contradicts the minimality of τ .

Finally, it is clear that $\tau \in \mathcal{D}$ is the unique cell with $\text{int}(\sigma) \subset \text{int}(\tau)$, because distinct cells in a cell decomposition have disjoint interior. \square

Definition 4.8 (Cellular maps and cellular Markov partitions). Let \mathcal{D}' and \mathcal{D} be two cell decompositions of X , and $f: X \rightarrow X$ be a continuous map. We say that f is *cellular* for $(\mathcal{D}', \mathcal{D})$ if the following condition is satisfied:

If $\sigma \in \mathcal{D}'$ is arbitrary, then $f(\sigma)$ is a cell in \mathcal{D} and $f|_\sigma$ is a homeomorphism of σ onto $f(\sigma)$.

If f is cellular with respect to $(\mathcal{D}', \mathcal{D})$ and \mathcal{D}' is a refinement of \mathcal{D} , then the pair $(\mathcal{D}', \mathcal{D})$ is called a *cellular Markov partition* for f .

Lemma 4.9. *Let \mathcal{D}' and \mathcal{D} be cell decompositions of X , and $f: X \rightarrow X$ be a continuous map that is cellular for $(\mathcal{D}', \mathcal{D})$. Suppose that $\sigma' \in \mathcal{D}'$, $\tau \in \mathcal{D}$, and $\tau \subset f(\sigma')$.*

Then there exists $\tau' \in \mathcal{D}'$ with $\tau' \subset \sigma'$ and $f(\tau') = \tau$.

Proof. Note that $f|_{\sigma'}$ is a homeomorphism of σ' onto $\sigma = f(\sigma')$. Pick a point $q \in \text{int}(\tau)$. Then there exists a point $p \in \sigma'$ with $f(p) = q$, and a cell $\tau' \in \mathcal{D}'$ with $p \in \text{int}(\tau')$. Then $\text{int}(\tau')$ meets σ' and so $\tau' \subset \sigma'$ (Lemma 4.3 (ii)). Moreover, $f(\tau')$ is a cell in \mathcal{D} with $q = f(p) \in \text{int}(f(\tau'))$. It follows that $\tau = f(\tau')$. \square

Proposition 4.10. *Let \mathcal{D}' and \mathcal{D} be two cell decompositions of X , and $f: X \rightarrow X$ be a continuous map. If $(\mathcal{D}', \mathcal{D})$ is a cellular Markov partition for f , then there exist unique cell decompositions \mathcal{D}^n of X for $n \in \mathbb{N}_0$ such that*

- (i) $\mathcal{D}^0 = \mathcal{D}$, $\mathcal{D}^1 = \mathcal{D}'$, and \mathcal{D}^{n+1} is a refinement of \mathcal{D}^n for $n \in \mathbb{N}_0$,
- (ii) each pair $(\mathcal{D}^{n+1}, \mathcal{D}^n)$, $n \in \mathbb{N}_0$, is a cellular Markov partition for f .

Note that this implies that \mathcal{D}^{n+k} is a refinement of \mathcal{D}^n , and f^k is cellular with respect to $(\mathcal{D}^{n+k}, \mathcal{D}^n)$ for all $n, k \in \mathbb{N}_0$. So $(\mathcal{D}^{n+k}, \mathcal{D}^n)$ is a cellular Markov partition for f^k .

Proof. The cell decompositions \mathcal{D}^n are constructed inductively. Let $\mathcal{D}^0 = \mathcal{D}$ and $\mathcal{D}^1 = \mathcal{D}'$.

The idea for constructing the refinement \mathcal{D}^2 of \mathcal{D}^1 is very simple: we want to decompose a cell $\sigma \in \mathcal{D}^1$ into cells in a similar way, as the cell

$f(\sigma) \in \mathcal{D}^0$ is decomposed by the cells $\sigma' \subset f(\sigma)$ in \mathcal{D}^1 . Accordingly, we define the set \mathcal{D}^2 as

$$\mathcal{D}^2 = \{(f|_{\sigma'})^{-1}(\sigma) : \sigma, \sigma' \in \mathcal{D}^1 \text{ and } \sigma \subset f(\sigma')\}.$$

Then \mathcal{D}^2 consists of cells. Indeed, if $\sigma, \sigma' \in \mathcal{D}^1$ and $\sigma \subset f(\sigma')$, then $f|_{\sigma'}$ is a homeomorphism of σ' onto $f(\sigma') \in \mathcal{D}^0$. Hence the preimage $\lambda := (f|_{\sigma'})^{-1}(\sigma)$ of the cell σ under this homeomorphism is a cell. Note that the cell $\sigma = f(\lambda)$ is uniquely determined by λ , but σ' in general is not.

We now show that \mathcal{D}^2 is a cell decomposition of X by verifying the conditions (i)–(iv) of Definition 4.1.

Condition (i): Let $x \in X$ be arbitrary. Then there exists $\sigma' \in \mathcal{D}^1$ with $x \in \sigma'$. The set $f(\sigma')$ is a cell in \mathcal{D}^0 . Since \mathcal{D}^1 is a refinement of \mathcal{D}^0 , there exists a cell $\sigma \in \mathcal{D}^1$ with $f(x) \in \sigma \subset f(\sigma')$. Then $(f|_{\sigma'})^{-1}(\sigma)$ is a cell in \mathcal{D}^2 that contains x . It follows that the cells in \mathcal{D}^2 cover X .

Condition (ii): Let $\lambda_1, \lambda_2 \in \mathcal{D}^2$ be arbitrary, and assume that $\text{int}(\lambda_1) \cap \text{int}(\lambda_2) \neq \emptyset$. We have to show that $\lambda_1 = \lambda_2$. We have $\text{int}(f(\lambda_i)) = f(\text{int}(\lambda_i))$ for $i = 1, 2$. So $f(\lambda_1)$ and $f(\lambda_2)$ are cells in \mathcal{D}^1 with a common interior point. Hence $\sigma := f(\lambda_1) = f(\lambda_2) \in \mathcal{D}^1$.

By definition of \mathcal{D}^2 there exist cells $\sigma_i \in \mathcal{D}^1$ with $\sigma \subset f(\sigma_i)$ and $\lambda_i = (f|_{\sigma_i})^{-1}(\sigma)$ for $i = 1, 2$. Let τ be the minimal cell in \mathcal{D}^0 that contains σ . Then $\sigma \subset \tau \subset f(\sigma_1) \cap f(\sigma_2)$. By Lemma 4.9 there exist cells $\tilde{\sigma}_i \in \mathcal{D}^1$ with $\tilde{\sigma}_i \subset \sigma_i$ and $f(\tilde{\sigma}_i) = \tau$ for $i = 1, 2$.

By Lemma 4.7 we have $\text{int}(\sigma) \subset \text{int}(\tau)$. Applying the homeomorphism $(f|_{\sigma_i})^{-1}$ to both sets in this inclusion, we obtain $\text{int}(\lambda_i) \subset \text{int}(\tilde{\sigma}_i)$ for $i = 1, 2$. It follows that $\tilde{\sigma}_1$ and $\tilde{\sigma}_2$ are cells in \mathcal{D}^1 with a common interior point. Hence $\tilde{\sigma}_1 = \tilde{\sigma}_2$, and so

$$\lambda_1 = (f|_{\sigma_1})^{-1}(\sigma) = (f|_{\tilde{\sigma}_1})^{-1}(\sigma) = (f|_{\tilde{\sigma}_2})^{-1}(\sigma) = (f|_{\sigma_2})^{-1}(\sigma) = \lambda_2$$

as desired.

Condition (iii): Let $\lambda \in \mathcal{D}^2$ and $x \in \partial\lambda$ be arbitrary. Then there exist cells $\sigma, \sigma' \in \mathcal{D}^1$ with $\sigma \subset f(\sigma')$ and $\lambda = (f|_{\sigma'})^{-1}(\sigma)$. Moreover, $\sigma = f(\lambda)$ and so $f(x) \in \partial\sigma$. By definition of a cell decomposition there exists a cell $\tilde{\sigma} \in \mathcal{D}^1$ with $f(x) \in \tilde{\sigma} \subset \partial\sigma$. Then $\tilde{\lambda} = (f|_{\sigma'})^{-1}(\tilde{\sigma})$ is a cell in \mathcal{D}^2 with $x \in \tilde{\lambda} \subset \partial\lambda$. It follows that $\partial\lambda$ is a union of cells in \mathcal{D}^2 .

We have verified conditions (i)–(iii) in the definition of a cell decomposition. Before we prove the last condition (iv), we will first show that \mathcal{D}^2 has the required properties of a refinement.

Indeed, if $\lambda \in \mathcal{D}^2$ and $\sigma, \sigma' \in \mathcal{D}^1$ are such that $\sigma \subset f(\sigma')$ and $\lambda = (f|_{\sigma'})^{-1}(\sigma)$, then

$$\lambda \subset (f|_{\sigma'})^{-1}(f(\sigma')) = \sigma'.$$

So every cell in \mathcal{D}^2 is contained in a cell in \mathcal{D}^1 .

Moreover, let $\sigma' \in \mathcal{D}^1$ and $x \in \sigma'$ be arbitrary. Then $f(\sigma')$ is a cell in \mathcal{D}^0 containing $f(x)$. Since \mathcal{D}^1 is a refinement of \mathcal{D}^0 there exists $\sigma \in \mathcal{D}^1$ with $f(x) \in \sigma \subset f(\sigma')$. Then $\lambda = (f|_{\sigma'})^{-1}(\sigma)$ is a cell in \mathcal{D}^2 with $x \in \lambda \subset \sigma'$. It follows that every cell in \mathcal{D}^1 is a union of cells in \mathcal{D}^2 .

Moreover, this is a finite union. Indeed, if $\sigma' \in \mathcal{D}^1$, then $f|_{\sigma'}$ induces a bijection of the cells in \mathcal{D}^2 contained in σ' and the cells in \mathcal{D}^1 contained in $f(\sigma') \in \mathcal{D}^0$. Since the latter set is finite, the former is finite as well.

Condition (iv): This can now easily be established. If $p \in X$ is arbitrary, then there exists a neighborhood U of p that meets only finitely many cells in \mathcal{D}^1 . Every cell in \mathcal{D}^2 that meets U must be contained in one of these finitely many cells from \mathcal{D}^1 . Since every cell in \mathcal{D}^1 contains only finitely many cells in \mathcal{D}^2 , it follows that U meets only finitely many cells in \mathcal{D}^2 .

We have proved that \mathcal{D}^2 is cell decomposition of X that is a refinement of \mathcal{D}^1 . It immediately follows from the definition of \mathcal{D}^2 that f is cellular with respect to $(\mathcal{D}^2, \mathcal{D}^1)$. Therefore, $(\mathcal{D}^2, \mathcal{D}^1)$ is a cellular Markov partition for f .

To show uniqueness of \mathcal{D}^2 suppose that $\tilde{\mathcal{D}}^2$ is another cell decomposition of X such that $(\tilde{\mathcal{D}}^2, \mathcal{D}^1)$ is a cellular Markov partition for f .

Then $\tilde{\mathcal{D}}^2 \subset \mathcal{D}^2$. Indeed, let $\lambda \in \tilde{\mathcal{D}}^2$ be arbitrary. Since $\tilde{\mathcal{D}}^2$ is a refinement of \mathcal{D}^1 , there exists a cell $\sigma' \in \mathcal{D}^1$ with $\lambda \subset \sigma'$. Moreover, $\sigma = f(\lambda)$ is a cell in \mathcal{D}^1 and $\sigma \subset f(\sigma')$. Since $f|_{\sigma'}$ is a homeomorphism of σ' onto $f(\sigma')$ it follows that

$$\lambda = (f|_{\sigma'})^{-1}(\sigma) \in \mathcal{D}^2.$$

If the inclusion $\tilde{\mathcal{D}}^2 \subset \mathcal{D}^2$ were strict, then there would be a cell in \mathcal{D}^2 whose interior would be disjoint from the interior of all the cells in $\tilde{\mathcal{D}}^2$. This is impossible, because $\tilde{\mathcal{D}}$ is a cell decomposition of X and so the interiors of the cells in $\tilde{\mathcal{D}}$ form a cover of X . So $\tilde{\mathcal{D}}^2 = \mathcal{D}^2$.

We have shown the existence and uniqueness of a cell decomposition \mathcal{D}^2 of X with the desired properties. Now \mathcal{D}^3 is constructed from $(\mathcal{D}^2, \mathcal{D}^1)$ in the same way as \mathcal{D}^2 was constructed from $(\mathcal{D}^1, \mathcal{D}^0)$. Continuing in this manner we get the desired existence and uniqueness of the cell decompositions \mathcal{D}^n . \square

Remark 4.11. The main idea of the previous proof can be summarized by saying that if the cell decompositions \mathcal{D}^n and \mathcal{D}^{n-1} have already been defined for some $n \in \mathbb{N}$, then one obtains the elements in \mathcal{D}^{n+1} by subdividing the cells $\sigma \in \mathcal{D}^n$ in the same way as the images $f(\sigma)$ in \mathcal{D}^{n-1} are subdivided by the cells in \mathcal{D}^n .

From this description it is clear that the “combinatorics” of the cells in the sequence \mathcal{D}^n , $n \in \mathbb{N}_0$, that is, their inclusion and intersection pattern, is determined by the pair $(\mathcal{D}^1, \mathcal{D}^0)$ and by the assignment $c \in \mathcal{D}^1 \mapsto f(c) \in \mathcal{D}^0$. Such an assignment of a cell in \mathcal{D}^0 to each cell in \mathcal{D}^1 is related to the concept of a “labeling” (see Definition 12.1). So for the combinatorics of the decompositions \mathcal{D}^n the only relevant information on the map f is its induced “labeling” $c \in \mathcal{D}^1 \mapsto f(c) \in \mathcal{D}^0$. It is not hard, but somewhat tedious, to formulate a precise statement based on a suitable notion of “combinatorial equivalence” for such sequences of cell decompositions (see the related Definition 12.2 where we define the notion of an isomorphism between cell complexes). We will not do this, because it would not add anything of substance, but content ourselves with the intuitive statement that the “combinatorics” of the sequence \mathcal{D}^n , $n \in \mathbb{N}_0$, is determined by the pair $(\mathcal{D}^1, \mathcal{D}^0)$, and the assignment $c \in \mathcal{D}^1 \mapsto f(c) \in \mathcal{D}^0$.

5. CELL DECOMPOSITIONS OF 2-SPHERES

In this section we study cell decompositions of 2-spheres and their relation to postcritically-finite branched covering maps. We first review some standard concepts and results from plane topology (see [Moi] for more details).

Let S^2 be a 2-sphere. An *arc* α in S^2 a homeomorphic image of the unit interval $[0, 1]$. The points corresponding corresponding to 0 and 1 under such a homeomorphism are called the *endpoints* of α . They are the unique points $p \in \alpha$ such that $\alpha \setminus \{p\}$ is connected. If p is an *interior point* of α , i.e., a point in α distinct from the endpoints, then there exist arbitrarily small open neighborhoods W of p such that $W \setminus \alpha$ has precisely two open connected components U and V .

A *closed Jordan region* X in S^2 is a homeomorphic image of the closed unit disk \mathbb{D} . The boundary ∂X of a closed Jordan region $X \subset S^2$ is a *Jordan curve*, i.e., the homeomorphic image of the unit circle $\partial \mathbb{D}$. If $J \subset S^2$ is a Jordan curve, then by the Schönflies Theorem there exists a homeomorphism $\varphi: S^2 \rightarrow \widehat{\mathbb{C}}$ such that $\varphi(J) = \partial \mathbb{D}$. In particular, the set $S^2 \setminus J$ has two connected component, both homeomorphic to \mathbb{D} . Note that arcs and closed Jordan region are cells of dimension 1 and 2, respectively.

Let \mathcal{D} be a cell decomposition of S^2 . Since the topological dimension of S^2 is equal to 2, no cell in \mathcal{D} can have dimension > 2 . We call the 2-dimensional cells in \mathcal{D} the *tiles*, and the 1-dimensional cells in \mathcal{D} the *edges* of \mathcal{D} . The *vertices* of \mathcal{D} are the points $v \in S^2$ such that $\{v\}$ is a cell in \mathcal{D} of dimension 0. So there is a somewhat subtle distinction

between vertices and cells of dimension 0: a vertex is an element of S^2 , while a cell of dimension 0 is a subset of S^2 with one element.

If c is a cell in \mathcal{D} , we denote by ∂c the boundary and by $\text{int}(c)$ the interior of c as introduced in the beginning of Section 4. Note that for edges and 0-cells c this is different from the boundary and the interior of c as a subset of the topological space S^2 .

We always assume that the sphere S^2 is *oriented*, i.e., one of the two generators of the singular homology group $H_2(S^2) \cong \mathbb{Z}$ (with coefficients in \mathbb{Z}) has been chosen as the *fundamental class* of S^2 . The orientation on S^2 induces an orientation on every Jordan region $X \subset S^2$ which in turn induces an orientation on ∂X and on every arc $\alpha \subset \partial X$.

This can be made precise by considering the fundamental homology classes representing orientations. For example, if X is a closed Jordan region in S^2 , then the fundamental class of S^2 maps to a generator of $H_2(X, \partial X)$ under the natural isomorphism

$$H_2(S^2) \cong H_2(S^2, S^2 \setminus \text{int}(X)) \cong H_2(X, \partial X) \cong \mathbb{Z}$$

induced by the inclusion map and excision. Hence we get an induced orientation on X .

On a more intuitive level, an orientation of an arc is just a selection of one of the endpoints as the *initial point* and the other endpoint as the *terminal point*. Let $X \subset S^2$ be a Jordan region in the oriented 2-sphere S^2 equipped with the induced orientation. If $\alpha \subset \partial X$ is an arc with a given orientation, then we say that X lies *to the left* or *to the right* of α depending on whether the orientation on α induced by the orientation of X agrees with the given orientation on α or not. Similarly, we say that with a given orientation of ∂X the Jordan region X lies to the left or right of ∂X .

To describe orientations, it is useful to introduce the notion of a flag. By definition a *flag* in S^2 is a triple (c_0, c_1, c_2) , where c_i is an i -dimensional cell for $i = 0, 1, 2$, $c_0 \subset \partial c_1$, and $c_1 \subset \partial c_2$. So a flag in S^2 is a closed Jordan region c_2 with an arc c_1 contained in its boundary, where the point in c_0 is a distinguished endpoint of c_1 . We orient the arc c_1 so that the point in c_0 is the initial point in c_1 . The flag is called *positively-* or *negatively-oriented* (for the given orientation on S^2) depending on whether c_2 lies to the left or to the right of the oriented arc c_1 .

A positively-oriented flag determines the orientation on S^2 uniquely. The standard orientation on $\widehat{\mathbb{C}}$ is the one for which the *standard flag* (c'_0, c'_1, c'_2) is positively-oriented, where $c'_0 = \{0\}$, $c'_1 = [0, 1] \subset \mathbb{R}$, and

$$c'_2 = \{z \in \mathbb{C} : 0 \leq \text{Re}(z) \leq 1, 0 \leq \text{Im}(z) \leq \text{Re}(z)\}.$$

Since edges and tiles in a cell decomposition \mathcal{D} of S^2 are arcs and closed Jordan regions, respectively, it makes sense to speak of oriented edges and tiles in \mathcal{D} . A *flag* in \mathcal{D} is a flag (c_0, c_1, c_2) , where c_0, c_1, c_2 are cells in \mathcal{D} . If c_i are i -dimensional cells in \mathcal{D} for $i = 0, 1, 2$, then (c_0, c_1, c_2) is a flag in \mathcal{D} if and only if $c_0 \subset c_1 \subset c_2$.

Cell decompositions of S^2 have additional properties that we summarize in the next lemma.

Lemma 5.1. *Let \mathcal{D} be a cell decomposition of S^2 . Then it has the following properties:*

- (i) *There are only finitely many cells in \mathcal{D} .*
- (ii) *The tiles in \mathcal{D} cover S^2 .*
- (iii) *Let X be a tile in \mathcal{D} . Then there exists a number $k \in \mathbb{N}$, $k \geq 2$, such that X contains precisely k edges e_1, \dots, e_k and k vertices v_1, \dots, v_k in \mathcal{D} . Moreover, these edges and vertices lie on the boundary ∂X of X , and we have*

$$\partial X = e_1 \cup \dots \cup e_k.$$

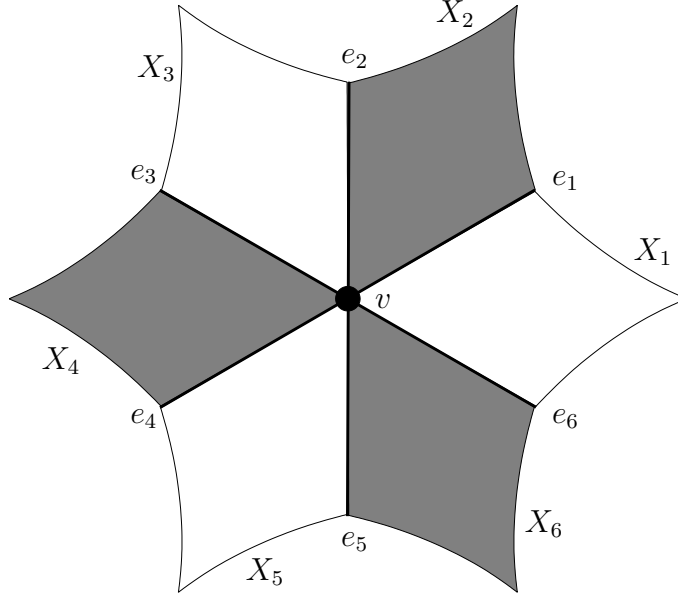
The indexing of these vertices and edges can be chosen such that $v_j \in \partial e_j \cap \partial e_{j+1}$ for $j = 1, \dots, k$ (where $e_{k+1} := e_1$).

- (iv) *Every edge $e \in \mathcal{D}$ is contained in the boundary of precisely two tiles \mathcal{D} . If X and Y are these tiles, then $\text{int}(X) \cup \text{int}(e) \cup \text{int}(Y)$ is an open set.*
- (v) *Let v be a vertex of \mathcal{D} . Then there exists a number $d \in \mathbb{N}$, $d \geq 2$, such that v is contained in precisely d tiles X_1, \dots, X_d , and d edges e_1, \dots, e_d in \mathcal{D} . We have $v \in \partial X_j$ and $v \in \partial e_j$ for each $j = 1, \dots, d$. Moreover, the indexing of these tiles and edges can be chosen such that $e_j \subset \partial X_j \cap \partial X_{j+1}$ for $j = 1, \dots, d$ (where $X_{d+1} := X_1$).*
- (vi) *The 1-skeleton of \mathcal{D} is connected and equal to the union of all edges in \mathcal{D} .*

Note that property (iii) actually holds for all tiles (i.e., 2-dimensional cells) in every cell decomposition. If the boundary of a tile X is subdivided into vertices and edges as in (iii), we say that X is a (*topological*) k -gon.

If the edge e and the tiles X and Y are as in (iv), then there exists a unique orientation of e such that X lies to the left and Y to the right of e .

We say that the cells $\{v\}, e_1, \dots, e_d, X_1, \dots, X_d$ as in (v) form the *cycle* of the vertex v and call d the *length* of the cycle. We refer to X_1, \dots, X_d as the tiles and to e_1, \dots, e_d as the edges of the cycle.

FIGURE 2. Cycle of a vertex v .

Proof. (i) By Definition 4.1 (iv) this follows from the compactness of S^2 and the fact that every point in S^2 has a neighborhood that meets only finitely many cells in \mathcal{D} .

(ii) The set consisting of all vertices and the union of all edges has empty interior (in the topological sense) by (i) and Baire's theorem. Hence the union of all tiles is a dense set in S^2 . Since this union is also closed by (i), it is all of S^2 .

(iii) Let X be a tile in \mathcal{D} . Then $\text{int}(X)$ does not meet any edge or vertex, and ∂X is a union of edges and vertices. Since there are only finitely many vertices, ∂X must contain an edge, and hence at least two vertices. Suppose v_1, \dots, v_k , $k \geq 2$, are all the vertices on ∂X . Since ∂X is a Jordan curve, we can choose the indexing of these vertices so that ∂X is a union of arcs α_j with pairwise disjoint interior such that α_j has the endpoints v_j and v_{j+1} for $j = 1, \dots, k$, where $v_{k+1} = v_1$. Then for each $j = 1, \dots, k$ the set $\text{int}(\alpha_j)$ is connected and lies in the 1-skeleton of the cell decomposition \mathcal{D} , it is disjoint from the 0-skeleton and has boundary contained in the 0-skeleton. It follows from Lemma 4.4 and Lemma 4.5 that there exists an edge e_j in \mathcal{D} with $\text{int}(e_j) = \text{int}(\alpha_j)$. Hence $\alpha_j = e_j$, and so α_j is an edge in \mathcal{D} . It is clear that ∂X does not contain other edges in \mathcal{D} . The statement follows.

(iv) Let e be an edge in \mathcal{D} . Pick $p \in \text{int}(e)$. By (ii) the point p is contained in some tile X in \mathcal{D} . By Lemma 4.3 we have $e \subset X$. On the other hand, $\text{int}(X)$ is disjoint from each edge and so $e \subset \partial X$. It

follows from the Schönflies theorem that the set X does not contain a neighborhood of p . Hence every neighborhood of p must meet tiles distinct from X . Since there are only finitely many tiles, it follows that there exists a tile Y distinct from X with $p \in Y$. By the same reasoning as before, we have $e \subset \partial Y$.

Let $q \in \text{int}(e)$ be arbitrary. Then there exist arbitrarily small open neighborhoods W of q such that $W \setminus \text{int}(e)$ consists of two connected components U and V . If W is small enough, then U and V do not meet ∂X . Since $q \in \overline{\text{int}(X)}$, one of the sets, say U , meets $\text{int}(X)$, and so $U \subset \text{int}(X)$. We can also assume that the set W is small enough so that it does not meet ∂Y either. By the same reasoning, U or V must be contained in $\text{int}(Y)$, and, since $\text{int}(X) \cap \text{int}(Y) = \emptyset$, we have $V \subset \text{int}(Y)$. Hence $\text{int}(X) \cup \text{int}(e) \cup \text{int}(Y)$ is a neighborhood of each point in $\text{int}(e)$ which implies that this set is open.

Suppose that Z is another tile in \mathcal{D} with $e \subset \partial Z$. Since $X \cup Y$ contains an open neighborhood for p , there exists a point $x \in \text{int}(Z)$ near p with $x \in X \cup Y$, say $x \in X$. Since the interior of a tile is disjoint from all other cells, we conclude $X = Z$. This shows the uniqueness of X and Y .

(v) Let v be a vertex of \mathcal{D} . If an edge e in \mathcal{D} contains v , then v is an endpoint of e and we orient e so that v is the initial point of e . By (ii) there exists a tile X_1 in \mathcal{D} with $v \in X_1$. Then $v \in \partial X_1$, and so by (iii) there exist two edges in ∂X_1 that contain v . For one of these oriented edges, denote it by e_1 , the tile X_1 will lie on the right of e_1 . Then $v \in e_1 \subset \partial X_1$ and X_1 will lie on the left of the other oriented edge. By (iv) there exists a unique tile $X_2 \neq X_1$ with $e_1 \subset \partial X_2$. Then X_2 will lie on the left of e_1 . By (iii) there exists a unique edge $e_2 \subset \partial X_2$ distinct from e_1 with $v \in e_2$. The tile X_2 will lie on the right of e_2 . We can continue in this manner to obtain tiles X_1, X_2, \dots and edges e_1, e_2, \dots that contain v and satisfy $X_j \neq X_{j+1}$, $e_j \neq e_{j+1}$, and $e_j \subset \partial X_j \cap \partial X_{j+1}$ for all $j \in \mathbb{N}$. Moreover, X_j will lie on the right and X_{j+1} on the left of the oriented edge e_j . Since there are only finitely many tiles, there exists a smallest number $d \in \mathbb{N}$ such that the tiles X_1, \dots, X_d are all distinct and X_{d+1} is equal to one of the tiles X_1, \dots, X_d . Since $X_1 \neq X_2$, we have $d \geq 2$.

Moreover, $X_{d+1} = X_1$. To see this we argue by contradiction and assume that X_{d+1} is equal to one of the tiles X_2, \dots, X_d say $X_{d+1} = X_j$. Note that $X_d \neq X_{d+1}$, so $2 \leq j \leq d-1$. Then $e = e_d$ is an edge with $v \in e$ that is contained in ∂X_d and in $\partial X_{d+1} = \partial X_j$. Hence $e = e_{j-1}$ or $e = e_j$. Since $X_{d+1} = X_j$ lies on the left of $e = e_d$, we must have

$e = e_{j-1}$. Then e is contained in the boundary of the three distinct tiles X_{j-1}, X_j, X_d which is impossible by (iv). So indeed $X_{d+1} = X_1$.

By a similar reasoning we can show that the edges e_1, \dots, e_d are all distinct. Indeed, suppose $e = e_j = e_k$, where $1 \leq j < k \leq d$. Then $k > j + 1$ and e is contained in the boundary of the three distinct tiles X_j, X_{j+1}, X_k which is again absurd.

To show that there are no other edges and tiles containing v note that by (iii) the set

$$U = \text{int}(X_1) \cup \text{int}(e_1) \cup \text{int}(X_2) \cup \dots \cup \text{int}(e_d) \cup \text{int}(X_{d+1})$$

is open. Moreover, its boundary ∂U consists of the point v and a closed set

$$A \subset \bigcup_{j=1}^d \partial X_j$$

disjoint from $\{v\}$. Hence v is an isolated boundary point of U which implies that $W = U \cup \{v\}$ is an open neighborhood of v .

If c is an arbitrary cell in \mathcal{D} with $v \in c$ and $c \neq \{v\}$, then $v \in \overline{\text{int}(c)}$. This implies that $\text{int}(c)$ meets U . Since interiors of distinct cells in \mathcal{D} are disjoint, this is only possible if c is equal to one of the edges e_1, \dots, e_d or one of the tiles X_1, \dots, X_d . The statement follows.

(vi) By (v) every vertex is contained in an edge. Hence the 1-skeleton E of \mathcal{D} is equal to the union of all edges in \mathcal{D} . To show that E is connected, let $x, y \in E$ be arbitrary. Since the tiles in \mathcal{D} cover S^2 , there exist tiles X and Y with $x \in X$ and $y \in Y$. The interior of each tile is disjoint from the 1-skeleton E , and so $x \in \partial X$ and $y \in \partial Y$. Since S^2 is connected, there exist tiles X_1, \dots, X_N in \mathcal{D} such that $X_1 = X$, $X_N = Y$, and $X_i \cap X_{i+1} \neq \emptyset$ for $i = 1, \dots, N-1$. The interior of a tile meets no other tile. Hence $\partial X_i \cap \partial X_{i+1} \neq \emptyset$ for $i = 1, \dots, N-1$. Since each set ∂X_i is connected, it follows that

$$K = \partial X_1 \cup \dots \cup \partial X_N$$

is a connected subset of E containing x and y . Hence E is connected. \square

Let $d \in \mathbb{N}$, $d \geq 2$, and the tiles X_j and edges e_j for $j \in \mathbb{N}$ be as defined in the proof of statement (v) of the previous lemma. Then we showed that $X_{d+1} = X_1$, but it is useful to point out that actually $X_j = X_{j+d}$ and $e_j = e_{j+d}$ for all $j \in \mathbb{N}$.

Indeed we have seen that $X_{d+1} = X_1$. Moreover, e_1, e_d, e_{d+1} are edges in \mathcal{D} that contain v and are contained in the boundary of the tile $X_1 = X_{d+1}$. Since there are only two such edges, $e_1 \neq e_d$, and $e_d \neq e_{d+1}$, we conclude that $e_{d+1} = e_1$. Then $e_1 = e_{d+1}$ is an edge

contained in the boundary of the tiles X_1, X_2, X_{d+2} . Since there are precisely two tiles containing an edge in its boundary, $X_1 \neq X_2$ and $X_1 = X_{d+1} \neq X_{d+2}$ it follows that $X_{d+2} = X_2$.

If we continue in this manner, shifting all indices by 1 in each step, we see that $e_{d+2} = e_2$, $X_{d+3} = X_3$, etc., as claimed.

Note that if we choose the indexing of the edges e_j and X_j as in the proof of statement (v) of the previous lemma, then for each $j \in \mathbb{N}$ the flag $(\{v\}, e_j, X_{j+1})$ in \mathcal{D} is positively-oriented, and flag $(\{v\}, e_j, X_j)$ is negatively-oriented. In other words, if e_j is oriented so that v is the initial point of e_j , then X_{j+1} lies to the left and X_j lies to the right of e_j .

Lemma 5.2. *Let \mathcal{D}' and \mathcal{D} be cell decompositions of S^2 , and $f: S^2 \rightarrow S^2$ be a cellular map for $(\mathcal{D}', \mathcal{D})$ such that $f|_X$ is orientation-preserving for each tile X in \mathcal{D}' .*

- (i) *Then f is a branched covering map on S^2 . Each critical point of f is a vertex of \mathcal{D}' .*
- (ii) *If in addition each vertex in \mathcal{D} is also a vertex in \mathcal{D}' , then every point in $\text{post}(f)$ is a vertex of \mathcal{D} . In particular, f is postcritically-finite, and hence a Thurston map if f is not a homeomorphism.*

Proof. (i) We will show that for each point $p \in S^2$, there exists $k \in \mathbb{N}$, an orientation-preserving homeomorphism φ of the open unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ onto a neighborhood W' of p , and an orientation-preserving homeomorphism ψ of a neighborhood $W \supset f(W')$ of $q = f(p)$ onto \mathbb{D} such that $\varphi(0) = p$, $\psi(q) = 0$, and

$$(\psi \circ f \circ \varphi)(z) = z^k$$

for all $z \in \mathbb{D}$. The desired relation between the points and maps can be represented by the commutative diagram

$$(5.1) \quad \begin{array}{ccc} p \in W' & \xrightarrow{f} & q \in W \\ \varphi \uparrow & & \downarrow \psi \\ 0 \in \mathbb{D} & \xrightarrow{z \mapsto z^k} & 0 \in \mathbb{D}. \end{array}$$

We will use the fact that if f is an orientation-preserving local homeomorphism near p , then we can take $k = 1$ and can always find suitable homeomorphisms φ and ψ .

Let $p \in S^2$ be arbitrary. Since S^2 is the disjoint union of the interior of the cells in \mathcal{D}' , the point p is contained in the interior of a tile or an edge in \mathcal{D}' , or is a vertex of \mathcal{D}' . Accordingly, we consider three cases.

Case 1. There exists a tile $X' \in \mathcal{D}'$ with $p \in \text{int}(X')$. Then $W' := \text{int}(X')$ is an open neighborhood of p , and $f|_{W'}$ is an orientation-preserving homeomorphism of $W' = \text{int}(X')$ onto $W := \text{int}(X)$, where $X = f(X') \in \mathcal{D}$. Hence f is a orientation-preserving local homeomorphism near p .

Case 2. There exists an edge $e' \in \mathcal{D}'$ with $p \in \text{int}(e')$. By Lemma 5.1 (iv) there exist distinct tiles $X', Y' \in \mathcal{D}'$ such that $e \subset \partial X' \cap \partial Y'$. Then $W' = \text{int}(X') \cup \text{int}(e') \cup \text{int}(Y')$ is an open neighborhood of p . Since f is cellular, $X = f(X')$ and $Y = f(Y')$ are tiles in \mathcal{D} and $e = f(e')$ is an edge in \mathcal{D} . Moreover, $e \subset \partial X \cap \partial Y$.

We orient e' so that X' lies to the left and Y' to the right of e' . Since f is orientation-preserving if restricted to cells in \mathcal{D}' , the tile X lies to the left, and Y to the right of the image e of e' . In particular, $X \neq Y$, and so the sets $\text{int}(X), \text{int}(e), \text{int}(Y)$ are pairwise disjoint, and their union is open. Since f is cellular and hence a homeomorphism if restricted to cells (and interior of cells), it follows that $f|_{W'}$ is a homeomorphism of W' onto the open set $W = \text{int}(X) \cup \text{int}(e) \cup \text{int}(Y)$. Moreover, it is clear that $f|_{W'}$ is orientation-preserving. Since W' is open and contains p , the map f is a orientation-preserving local homeomorphism near p .

Case 3. The point p is a vertex of \mathcal{D}' . As in the proof of Lemma 5.1 (v) we can choose tiles $X'_j \in \mathcal{D}'$ and edges $e'_j \in \mathcal{D}$ for $j \in \mathbb{N}$ that contain p and satisfy $X'_j \neq X'_{j+1}$, $e'_j \neq e'_{j+1}$, and $e'_j \subset \partial X'_j \cap \partial X'_{j+1}$ for all $j \in \mathbb{N}$. There exists $d' \in \mathbb{N}$ such that $X'_{d'+1} = X'_1$ and such that the tiles $X'_1, \dots, X'_{d'}$ and the edges $e'_1, \dots, e'_{d'}$ are all distinct and such that

$$W' = \{p\} \cup \text{int}(X'_1) \cup \text{int}(e'_1) \cup \text{int}(X'_2) \cup \dots \cup \text{int}(e'_{d'})$$

is an open neighborhood of p .

Moreover, by the remark following Lemma 5.1, we know that $X'_j = X'_{j+d'}$ and $e'_j = e'_{j+d'}$ for all $j \in \mathbb{N}$.

Define $X_j = f(X'_j)$ and $e_j = f(e'_j)$ for $j \in \mathbb{N}$. Since f is cellular for $(\mathcal{D}', \mathcal{D})$, the set X_j is a tile and e_j an edge in \mathcal{D} . Moreover, $e_j \subset \partial X_j \cap \partial X_{j+1}$ for $j \in \mathbb{N}$. Since X'_j and X'_{j+1} are distinct tiles containing the edge e'_{j+1} in their boundaries, it follows by an argument as in Case 2 above that $X_j \neq X_{j+1}$ for $j \in \mathbb{N}$. Moreover, since e'_j and e'_{j+1} are distinct edges in \mathcal{D}' contained in X'_{j+1} , and $f|_{X'_{j+1}}$ is a homeomorphism, we also have $e_j \neq e_{j+1}$ for $j \in \mathbb{N}$.

As in the proof of Lemma 5.1 (v) we see that there exists a number $d \in \mathbb{N}$, $d \geq 2$, such that $X_{d+1} = X_1$, and such the tiles X_1, \dots, X_d and the edges e_1, \dots, e_d are all distinct. Moreover,

$$W = \{q\} \cup \text{int}(X_1) \cup \text{int}(e_1) \cup \text{int}(X_2) \cup \dots \cup \text{int}(e_d)$$

is an open neighborhood of $q = f(p)$, and $X_i = X_{j+d}$ and $e_j = e_{j+d}$ for all $j \in \mathbb{N}$.

The periodicity properties of the indexing of the tiles X'_j and X_j imply that $d \leq d'$ and that d is a divisor of d' . Hence there exists $k \in \mathbb{N}$ such that $d' = kd$.

We now claim that after suitable coordinate changes near p and q , the map f can be given the form $z \mapsto z^k$.

For $N \in \mathbb{N}$, $N \geq 2$, and $j \in \mathbb{N}$ define half-open line segments

$$L_j^N = \{re^{2\pi i j/N} : 0 \leq r < 1\} \subset \mathbb{D}$$

and sectors

$$\Sigma_j^N = \{re^{it} : 2\pi(j-1)/N \leq t \leq 2\pi j/N \text{ and } 0 \leq r < 1\} \subset \mathbb{D}.$$

We then construct a homeomorphism $\psi: W \rightarrow \mathbb{D}$ with $\psi(q) = 0$ as follows. For each $j = 1, \dots, d$ we first map the half-open arc $\{q\} \cup \text{int}(e_j)$ homeomorphically to the half-open line segment L_j^d . Then q is mapped to 0, so these maps are consistently defined for q . Since X_j is a Jordan region, we can extend the homeomorphisms on $\{q\} \cup \text{int}(e_{j-1}) \subset \partial X_j$ and on $\{q\} \cup \text{int}(e_j) \subset \partial X_j$ to a homeomorphism of

$$\{q\} \cup \text{int}(e_{j-1}) \cup \text{int}(e_j) \cup \text{int}(X_j)$$

onto the sector Σ_j^d for each $j = 2, \dots, d+1$. Since the sets

$$\{q\}, \text{int}(e_1), \dots, \text{int}(e_d), \text{int}(X_2), \dots, \text{int}(X_{d+1}) = \text{int}(X_1)$$

are pairwise disjoint and have W as a union, these homeomorphisms paste together to a well-defined homeomorphism ψ of W onto \mathbb{D} . Note that $\psi(q) = 0$ and $\psi(X_j \cap W) = \Sigma_j^d$ for each $j = 1, \dots, d$.

A homeomorphism $\varphi: \mathbb{D} \rightarrow W'$ is defined as follows. If $z \in \mathbb{D}$ is arbitrary, then $z \in \Sigma_j^{d'}$ for some $j = 1, \dots, d'$. Hence $z^k \in \Sigma_j^d$, and so $\psi^{-1}(z^k) \in X_j \cap W$. Since f is a homeomorphism of $X'_j \cap W'$ onto $X_j \cap W$, it follows that $(f|X'_j)^{-1}(\psi^{-1}(z^k))$ is defined and lies in $X'_j \cap W'$.

We set

$$\varphi(z) = (f|X'_j)^{-1}(\psi^{-1}(z^k)).$$

It is straightforward to verify that φ is well-defined and a homeomorphism of \mathbb{D} onto W' with $\varphi(0) = p$. It follows from the definition of φ that $(\psi \circ f \circ \varphi)(z) = z^k$ for $z \in \mathbb{D}$, and so we have the diagram (5.1).

We assume that the tiles X'_j and the edges e'_j are indexed by the procedure in the proof of Lemma 5.1 (v). Then each flag $(\{p\}, e'_j, X'_{j+1})$ is positively-oriented (see the remark after the proof of Lemma 5.1). Since $f|X'_j$ is orientation-preserving, this implies that the flag $(\{q\}, e_j, X_{j+1})$ is also positively-oriented. We conclude that ψ is orientation-preserving, since ψ maps the positively-oriented flag $(\{q\}, e_j, X_{j+1})$ in S^2 to the

positively-oriented flag $(\{0\}, L_j^d, \Sigma_{j+1}^d)$ in $\widehat{\mathbb{C}}$. As follows from its definition, the map φ is then also orientation-preserving. Hence ψ and φ are local homeomorphisms as desired.

We have shown that in all cases the map f has a local behavior as claimed. It follows that f is a branched covering map. Moreover, we have seen that f near each point is a local homeomorphism unless p is a vertex of \mathcal{D}' . It follows that each critical point of f is a vertex of \mathcal{D}' .

(ii) Suppose in addition that every vertex of \mathcal{D} is also a vertex of \mathcal{D}' . Let p be a critical point of f . Then by (i) the point p is a vertex of \mathcal{D}' . Since f is cellular for $(\mathcal{D}', \mathcal{D})$, the point $f(p)$ is a vertex of \mathcal{D} . Hence $f(p)$ is also a vertex of \mathcal{D}' , and we can apply the argument again, to conclude that $f^2(p)$ is a vertex of \mathcal{D} , etc. It follows that $\text{post}(f)$ is a subset of the set of vertices of \mathcal{D} . In particular, $\text{post}(f)$ is finite, and so f is postcritically-finite. \square

Remark 5.3. Let the map $f: S^2 \rightarrow S^2$ and the cell decompositions \mathcal{D}' and \mathcal{D} be as in the previous lemma, and let p be a vertex in \mathcal{D}' . Then $q = f(p)$ is a vertex in \mathcal{D} . If d' and d are the lengths of the cycles of p in \mathcal{D}' and q in \mathcal{D} , respectively, then $d' = d \deg_f(p)$. Moreover, the tiles and edges of the cycle of q in \mathcal{D} are the images under f of the tiles and edges of the cycle of p in \mathcal{D} . This was established in Case 3 of the proof of Lemma 5.2.

Lemma 5.4. *Let $f: S^2 \rightarrow S^2$ be a branched covering map and \mathcal{D} a cell decomposition of S^2 such that every point in $f(\text{crit}(f))$ is a vertex in \mathcal{D} . Then there exists a unique cell decomposition \mathcal{D}' of S^2 such that f is cellular with respect to $(\mathcal{D}', \mathcal{D})$.*

Proof. To show existence we define \mathcal{D}' to be the set of all cells $c \subset S^2$ such that $f(c)$ is a cell in \mathcal{D} and $f|_c$ is a homeomorphism of c onto $f(c)$. It is clear that \mathcal{D}' does not contain cells of dimension > 2 . As usual we call the cells c in \mathcal{D}' edges or tiles depending on whether c has dimension 1 or 2, respectively. The vertices p of \mathcal{D}' are the points in S^2 such that $\{p\}$ is a cell in \mathcal{D}' of dimension 0.

It is clear that the set of vertices of \mathcal{D}' is equal to $f^{-1}(\mathbf{V})$, where \mathbf{V} is the set of vertices of \mathcal{D} .

To show that \mathcal{D}' is a cell decomposition of S^2 , we first establish two claims.

Claim 1. If $p \in S^2$ and $q = f(p) \in \text{int}(X)$ for some tile $X \in \mathcal{D}$, then there exists a unique tile $X' \in \mathcal{D}'$ with $p \in X'$.

In this case let $U = \text{int}(X)$. Then U is an open and simply connected set in the complement of $\mathbf{V} \supset f(\text{crit}(f))$. Hence there exists a unique continuous map $g: U \rightarrow U' := g(U)$ with $f \circ g = \text{id}_U$ and $g(q) = p$.

The map g is a homeomorphism onto its image U' . Hence $U' \subset S^2$ is open and simply connected.

We equip S^2 with some base metric inducing the standard topology. In the following metric terms will refer to this metric. Recall that $\mathcal{N}^\epsilon(A)$ denotes the open ϵ -neighborhood of a set $A \subset S^2$. Then f has the following property: for all $w \in S^2$ and all $\epsilon > 0$, there exists $\delta > 0$ such that

$$(5.2) \quad f^{-1}(B(w, \delta)) \subset \mathcal{N}^\epsilon(f^{-1}(w)).$$

Indeed, if for some $w \in S^2$ and $\epsilon > 0$ there is no such δ , then there exists a sequence $\{z_i\}$ in $S^2 \setminus \mathcal{N}^\epsilon(f^{-1}(w))$ such that $f(z_i) \in B(w, 1/i)$ for all $i \in \mathbb{N}$. By passing to a subsequence, we may assume that $z_i \rightarrow z \in S^2$. Then $f(z) = \lim_{i \rightarrow \infty} f(z_i) = w$, while

$$\text{dist}(z, f^{-1}(w)) = \lim_{i \rightarrow \infty} \text{dist}(z_i, f^{-1}(w)) \geq \epsilon.$$

This is a contradiction showing (5.2).

We want to prove that g has a continuous extension to $\overline{U} = X$. For this it suffices to show that $\{g(w_i)\}$ converges whenever $\{w_i\}$ is a sequence in U converging to a point $w \in \partial U$. Since g is a right inverse of f , it follows that the limit points of $\{g(w_n)\}$ are contained in $f^{-1}(w)$. Since f is finite-to-one, the point w has finitely many preimages z_1, \dots, z_m under f .

We can choose $\epsilon > 0$ so small that the sets $B(z_i, \epsilon)$, $i = 1, \dots, m$, are pairwise disjoint. By (5.2) we can find $\delta > 0$ such that

$$(5.3) \quad f^{-1}(B(w, \delta)) \subset \bigcup_{i=1}^m B(z_i, \epsilon).$$

The set $\overline{U} = X$ is a closed Jordan region, and hence locally connected. So there exists an open connected set $V \subset U$ such that \overline{V} is a neighborhood of w in \overline{U} and $\overline{V} \subset B(w, \delta)$. Then $g(V)$ is connected subset of $f^{-1}(B(w, \delta))$. Since the union on the right hand side of (5.3) is disjoint, the set $g(V)$ must be contained in one of the sets of this union, say $g(V) \subset B(z_k, \epsilon)$. Now $w_i \in V$ for sufficiently large i , and so all limit points of $\{g(w_i)\}$ are contained in $\overline{g(V)} \subset \overline{B(z_k, \epsilon)}$. On the other hand, the only possible limit points of $\{g(w_i)\}$ are z_1, \dots, z_m , and z_k is the only one contained in $\overline{B(z_k, \epsilon)}$. This implies $\{g(w_i)\} \rightarrow z_k$. So g has indeed a continuous extension to \overline{U} . We also denote it by g . It is clear that

$$(5.4) \quad f \circ g = \text{id}_{\overline{U}}.$$

This implies that g is a homeomorphism of $\overline{U} = X$ onto its image $X' := g(\overline{U}) = \overline{g(U)}$. Then X' is a closed Jordan region, and by (5.4) the map $f|_{X'}$ is a homeomorphism of X' onto $\overline{U} = X$. Hence X' is a tile in \mathcal{D}' with $p \in g(U) \subset X'$.

So a tile $X' \in \mathcal{D}'$ containing p exists. We want to show that it is the only tile in \mathcal{D}' containing p . Indeed, suppose $Y' \in \mathcal{D}'$ is another tile with $p \in Y'$. Then $f(Y')$ is a tile in \mathcal{D} containing the point $q = f(p) \in \text{int}(X)$. Hence $f(Y') = X$, and so $f|_{Y'}$ is a homeomorphism of Y' onto X . Let $h = (f|_X)^{-1}$. Then g and h are both inverse branches of f defined on the simply connected region U with $g(q) = p = h(q)$. Hence h and g agree on U , and so by continuity also on \overline{U} . It follows that $X' = g(X) = h(X) = Y'$ as desired.

Claim 2. If $p \in S^2$ and $q = f(p) \in \text{int}(e)$ for some edge $e \in \mathcal{D}$, then there exists a unique edge $e' \in \mathcal{D}'$, and precisely two distinct tiles X' and Y' in \mathcal{D}' that contain p . Moreover, $e' \subset \partial X' \cup \partial Y'$.

By Lemma 5.1 (iv) we know that there are precisely two distinct tiles $X, Y \in \mathcal{D}$ that contain e in their boundary, and that $U = \text{int}(X) \cup \text{int}(e) \cup \text{int}(Y)$ is an open and simply connected region in the complement of the set $\mathbf{V} \supset f(\text{crit}(f))$. Hence there exists a unique continuous map $g: U \rightarrow S^2$ with $g(q) = p$ and $f \circ g = \text{id}_U$. As before one can show that the maps $g_1 := g|_{\text{int}(X)}$ and $g_2 := g|_{\text{int}(Y)}$ have continuous extensions to X and Y , respectively. We use the same notation g_1 and g_2 for these extensions. It is clear that $g_1|_e = g_2|_e$. Moreover, g_1 is a homeomorphism of X onto a closed Jordan region $X' = g_1(X)$ with inverse map $f|_{X'}$. In particular, X' is a tile in \mathcal{D}' . Similarly, $Y' = g_2(Y)$ is a tile in \mathcal{D}' . The tiles X' and Y' are distinct, because f maps them to different tiles in \mathcal{D} . Moreover, $e' := g_1(e) = g_2(e)$ is an edge in \mathcal{D}' with $p \in e' \subset \partial X' \cap \partial Y'$.

It remains to prove the uniqueness part. If \tilde{e} is another edge in \mathcal{D}' with $p \in \tilde{e}$, then f is a homeomorphism of \tilde{e} onto e . Then $(f|_{\text{int}(e')})^{-1}$ and $f(\text{int}(\tilde{e}))^{-1}$ are right inverses of f defined on the open arc $\text{int}(e)$ that both map q to p . Hence these right inverses must agree on $\text{int}(e)$. By continuity this implies $(f|_{e'})^{-1} = (f|_{\tilde{e}})^{-1}$ on e , and so $e' = (f|_{e'})^{-1}(e) = (f|_{\tilde{e}})^{-1}(e) = \tilde{e}$.

If Z' is another tile in \mathcal{D}' with $p \in Z'$, then f maps $\partial Z'$ homeomorphically to the boundary $\partial f(Z')$ of the tile $f(Z') \in \mathcal{D}$. Moreover, $p \in \partial Z'$; for otherwise $f(p)$ would lie in the set $\text{int}(f(Z'))$ which is disjoint of e . It follows that there is an edge in \mathcal{D}' that contains p and is contained in the boundary of $\partial Z'$. Since this edge in \mathcal{D}' is unique, as we have just seen, we know that $e' \subset \partial Z'$. Note that $X' \cup Y' \supset g(U) \supset \text{int}(e')$, and so $X' \cup Y'$ contains an open neighborhood for each point in $\text{int}(e')$. Since

$e' \subset \partial Z'$ there exists a point $x \in \text{int}(Z')$ near p with $x \in X' \cup Y'$, say $x \in X'$. Then $f(x)$ is contained in the interior of the tile $f(Z') \in \mathcal{D}$. Since $x \in X' \cap Z'$ and X' and Z' are both tiles in \mathcal{D}' we conclude $X' = Z'$ by the first claim. This concludes the proof of Claim 2.

Now that we have established the claims, we can show that \mathcal{D} is a cell decomposition of S^2 by verifying conditions (i)–(iv) of Definition 4.1.

Condition (i): If $p \in S^2$ is arbitrary, then $f(p)$ is a vertex of \mathcal{D} or $f(p)$ lies in the interior of an edge or in the interior of a tile in \mathcal{D} . In the first case p is a vertex of \mathcal{D}' , and in the other two cases p lies in cells in \mathcal{D}' by Claim 1 and Claim 2. It follows that the cells in \mathcal{D}' cover S^2 .

Condition (ii): Let σ, τ be cells in \mathcal{D}' with $\text{int}(\sigma) \cap \text{int}(\tau) \neq \emptyset$. Then $f(\sigma)$ and $f(\tau)$ are cells in \mathcal{D} with $\text{int}(f(\sigma)) \cap \text{int}(f(\tau)) \neq \emptyset$. Hence $\lambda = f(\sigma) = f(\tau)$. In particular, σ and τ have the same dimension.

If σ and τ are both tiles, then $\sigma = \tau$ by Claim 1, because every point in $\text{int}(\sigma) \cap \text{int}(\tau) \neq \emptyset$ has an image under f in $\text{int}(\lambda)$. Similarly, if σ and τ are edges, then $\sigma = \tau$ by Claim 2.

If σ and τ consist of vertices in \mathcal{D}' , then the relation $\text{int}(\sigma) \cap \text{int}(\tau) \neq \emptyset$ trivially implies $\sigma = \tau$.

Condition (iii): Let $c' \in \mathcal{D}'$ be arbitrary. Then $f|_{c'}$ is a homeomorphism of c' onto the cell $c = f(c') \in \mathcal{D}$. Note that $(f|_{c'})^{-1}(\sigma) \in \mathcal{D}'$ whenever $\sigma \in \mathcal{D}$ and $\sigma \subset c$. Since $\partial c' = (f|_{c'})^{-1}(\partial c)$ and ∂c is a union of cells in \mathcal{D} , it follows that $\partial c'$ is a union of cells in \mathcal{D}' .

Condition (iv): To establish the final property of a cell decomposition for \mathcal{D}' , we will show that \mathcal{D}' consists of only finitely many cells. Indeed, let $N_i \in \mathbb{N}$ be the number of cells of dimension i in \mathcal{D} for $i = 0, 1, 2$. Since the vertices in \mathcal{D}' are the preimages of the vertices of \mathcal{D} , we have at most $\deg(f)N_0$ vertices in \mathcal{D}' .

Pick one point in the interior of each edge in \mathcal{D} . The set M of these points consists of N_1 elements. If $q \in M$, then $q \notin \mathbf{V} \supset f(\text{crit}(f))$, and so q is not a critical value of f . Hence $\#f^{-1}(M) = N_1 \deg(f)$. It follows from Claim 2 that each element of $f^{-1}(M)$ is contained in a unique edge in \mathcal{D}' , and it follows from the definition of \mathcal{D} that each edge in \mathcal{D}' contains a unique point in $f^{-1}(M)$. Hence the number of edges in \mathcal{D}' is equal to $\#f^{-1}(M) = N_1 \deg(f)$.

Similarly, pick a point in the interior of each tile in \mathcal{D} and let M be the set of these points. Then $\#f^{-1}(M) = N_2 \deg(f)$ and by the same reasoning as above based on Claim 1, we see that the number of tiles in \mathcal{D}' is equal to $\#f^{-1}(M) = N_2 \deg(f)$.

We have shown that \mathcal{D}' is a cell decomposition of S^2 . It follows immediately from the definition of \mathcal{D}' that f is cellular with respect to $(\mathcal{D}', \mathcal{D})$.

To show uniqueness of \mathcal{D}' suppose that $\tilde{\mathcal{D}}$ is another cell decomposition such that f is cellular with respect to $(\tilde{\mathcal{D}}, \mathcal{D})$. Then by definition of \mathcal{D}' every cell in $\tilde{\mathcal{D}}$ also lies in \mathcal{D}' . So we have $\tilde{\mathcal{D}} \subset \mathcal{D}'$. If this inclusion were strict, then there would be a cell $c \in \mathcal{D}'$ whose interior $\text{int}(c) \neq \emptyset$ is disjoint from the interior of all cells in $\tilde{\mathcal{D}}$. This is impossible, since these interiors form a cover of S^2 . Hence $\tilde{\mathcal{D}} = \mathcal{D}'$. \square

Remark 5.5. No Thurston maps $f: S^2 \rightarrow S^2$ with $\#\text{post}(f) \in \{0, 1\}$ exist. Indeed, suppose that f is such a map. Then $U := S^2 \setminus \text{post}(f)$ is simply connected, and so there exists a continuous map $g: U \rightarrow S^2$ with $f \circ g = \text{id}_U$.

If $\#\text{post}(f) = 0$, we have $U = S^2$ and so we conclude that g is a homeomorphism onto its image. This image must be all of S^2 . Hence g and f are homeomorphisms, contradicting our assumption $\deg(f) \geq 2$ (see Definition 3.1).

If $\#\text{post}(f) = 1$, we have $U = S^2 \setminus \{p\}$ for some $p \in S^2$. Then by an argument as in the proof of Claim 1 in Lemma 5.4, one can show that g has a continuous extension to the point p , and hence to S^2 . If we denote this extension to S^2 also by g , then $f \circ g = \text{id}_{S^2}$, and again we conclude that f is a homeomorphism and obtain a contradiction.

6. CELL DECOMPOSITIONS INDUCED BY THURSTON MAPS

Let $f: S^2 \rightarrow S^2$ be a Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve such that $\text{post}(f) \subset \mathcal{C}$. In this section we will show that the pair (f, \mathcal{C}) induces natural cell decompositions of S^2 .

By the Schönflies theorem there are two closed Jordan regions $X_{\mathbf{b}}^0, X_{\mathbf{w}}^0 \subset S^2$ whose boundary is \mathcal{C} . Our notation for these regions is suggested by the fact that we often think of $X_{\mathbf{b}}^0$ as being assigned or carrying the color “black”, represented by the symbol \mathbf{b} , and $X_{\mathbf{w}}^0$ as being colored “white” represented by \mathbf{w} . We will discuss this more precisely later in this section (see Lemma 6.2).

The sets $X_{\mathbf{b}}^0$ and $X_{\mathbf{w}}^0$ are topological cells of dimension 2. We call them *tiles of order 0* or *0-tiles*. The postcritical points of f are on the boundary of $X_{\mathbf{w}}^0$ and $X_{\mathbf{b}}^0$. We consider them as *vertices* of $X_{\mathbf{w}}^0$ and $X_{\mathbf{b}}^0$, and the closed arcs of \mathcal{C} between vertices as the *edges* of the 0-tiles. In this way, we think of $X_{\mathbf{w}}^0$ and $X_{\mathbf{b}}^0$ as topological m -gons where $m = \#\text{post}(f) \geq 2$ (see Remark 5.5). To emphasize that these edges

and vertices belong to 0-tiles, we call them *0-edges* and *0-vertices*. A *0-cell* is a 0-tile, a 0-edge, or a set consisting of a 0-vertex. Obviously, the 0-cells form a cell decomposition of S^2 that we denote by $\mathcal{D}^0 = \mathcal{D}^0(f, \mathcal{C})$.

Since every point in $\text{post}(f)$ is a vertex of \mathcal{D}^0 , we can apply Lemma 5.4 to obtain a unique cell decomposition $\mathcal{D}^1 = \mathcal{D}^1(f, \mathcal{C})$ such that f is cellular with respect to $(\mathcal{D}^1, \mathcal{D}^0)$. The vertices of \mathcal{D}^1 are precisely the points whose image is a vertex of \mathcal{D}^0 . In particular, since $f(\text{post}(f)) \subset \text{post}(f)$, or equivalently $\text{post}(f) \subset f^{-1}(\text{post}(f))$, it follows that every point in $\text{post}(f)$ is a vertex for \mathcal{D}^1 . Hence we can apply Lemma 5.4 again and obtain a cell decomposition $\mathcal{D}^2 = \mathcal{D}^2(f, \mathcal{C})$ such that f is cellular with respect to $(\mathcal{D}^2, \mathcal{D}^1)$. Continuing in this manner, we obtain cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ of S^2 for $n \in \mathbb{N}_0$ such that f is cellular for $(\mathcal{D}^{n+1}, \mathcal{D}^n)$ for all $n \in \mathbb{N}_0$.

We call the elements in \mathcal{D}^n the *n-cells* for (f, \mathcal{C}) , or simply *n-cells* if f and \mathcal{C} are understood. We call n the *order* of an *n-cell*. When we speak of *n-cells*, then n always refers to this order and not to the dimension of the cell. An *n-cell* of dimension 2 is called an *n-tile*, and an *n-cell* of dimension 1 an *n-edge*. An *n-vertex* is a point $p \in S^2$ such that $\{p\}$ is an *n-cell* of dimension 0. With f and \mathcal{C} understood we denote the set of all *n-tiles*, *n-edges*, and *n-vertices* by \mathbf{X}^n , \mathbf{E}^n , and \mathbf{V}^n , respectively.

In the following proposition we summarize properties of the cell decompositions \mathcal{D}^n .

Proposition 6.1. *Let $k, n \in \mathbb{N}_0$, let $f: S^2 \rightarrow S^2$ be a Thurston map, $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$, $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$, and $m = \# \text{post}(f)$.*

- (i) *The map f^k is cellular with respect to $(\mathcal{D}^{n+k}, \mathcal{D}^n)$. In particular, if τ is any $(n+k)$ -cell, then $f^k(\tau)$ is an n -cell, and $f^k|_\tau$ is a homeomorphism of τ onto $f(\tau)$.*
- (ii) *Let σ be an n -cell. Then $f^{-k}(\sigma)$ is equal to the union of all $(n+k)$ -cells τ with $f^k(\tau) = \sigma$.*
- (iii) *The 0-skeleton of \mathcal{D}^n is the set $\mathbf{V}^n = f^{-n}(\text{post}(f))$, and we have $\mathbf{V}^n \subset \mathbf{V}^{n+k}$. The 1-skeleton of \mathcal{D}^n is equal to $f^{-n}(\mathcal{C})$.*
- (iv) *We have $\#\mathbf{V}^n \leq m \deg(f)^n$, $\#\mathbf{E}^n = m \deg(f)^n$, and $\#\mathbf{X}^n = 2 \deg(f)^n$ for $n \in \mathbb{N}_0$.*
- (v) *The n -edges are precisely the closures of the connected components of $f^{-n}(\mathcal{C}) \setminus f^{-n}(\text{post}(f))$. The n -tiles are precisely the closures of the connected components of $S^2 \setminus f^{-n}(\mathcal{C})$.*
- (vi) *Every n -tile is an m -gon, i.e., the number of n -edges and n -vertices contained in its boundary is equal to m .*

Proof. (i) This immediately follows from the facts that f is cellular with respect to $(\mathcal{D}^{n+1}, \mathcal{D}^n)$ for each n , and that compositions of cellular maps are cellular (if, as in our case, the obvious compatibility requirement for the cell decompositions involved is satisfied).

(ii) It follows from (i) and Lemma 5.4 that \mathcal{D}^{n+k} is the unique cell decomposition of S^2 such that f^k is cellular with respect to $(\mathcal{D}^{n+k}, \mathcal{D}^n)$. Moreover, recall from the proof of Lemma 5.4 that a topological cell $c \subset S^2$ is an $(n+k)$ -cell if and only if $f^k(c)$ is an n -cell and $f^k|_c$ is a homeomorphism of c onto $f^k(c)$.

This immediately implies the statement if $\sigma = \{q\}$, where q is an n -vertex.

Suppose σ is equal to an n -edge e . Let M be the union of all $(n+k)$ -edges e' with $f^k(e') = e$. It is clear that $M \subset f^{-k}(e)$. To see the converse inclusion, let $p \in f^{-k}(e)$ be arbitrary.

If $p \in f^{-k}(\text{int}(e))$, then from Claim 2 in the proof of Lemma 5.4 it follows that there exists an $(n+k)$ -edge e' with $p \in e'$. Then $f^k(e')$ is an n -edge that contains $q = f^k(p) \in \text{int}(e)$. Hence $e = f^k(e')$, and so $f^{-k}(\text{int}(e)) \subset M$.

If $p \in f^{-k}(\partial e)$, then $q = f^k(p) \in \partial e$ is an n -vertex, and so p is an $(n+k)$ -vertex as we have seen. It follows from Case 3 in the proof of Lemma 5.2 that there exists an $(n+k)$ -edge e' that contains p with $f^k(e') = e$. Hence $p \in M$. We conclude that $f^{-k}(e) = f^{-k}(\text{int}(e)) \cup f^{-k}(\partial e) \subset M$, and so $M = f^{-k}(e)$ as desired.

If σ is equal to an n -tile X , let M be the union of all $(n+k)$ -tiles X' with $f^k(X') = X$. Then $M \subset f^{-k}(X)$. For the converse inclusion let again $p \in f^{-k}(X)$ be arbitrary.

If $p \in f^{-k}(\text{int}(X))$, then by Claim 1 in the proof of Lemma 5.4 there exists an $(n+k)$ -tile with $p \in X'$. Similarly as above we conclude $f^k(X') = X$, and so $p \in M$.

If $p \in f^{-k}(\partial X)$, then $q = f^k(p) \in \partial X$, and so there exists an n -edge $e \subset \partial X$ with $q \in e$. By what we have seen in the first part of the proof, there exists an $(n+k)$ -edge e' with $f^k(e') = e$ and $p \in e'$. By Lemma 5.1 (iv) there exist two $(n+k)$ -tiles X' and Y' with $e' \subset \partial X' \cap \partial Y'$. Then $f^k(X')$ and $f^k(Y')$ are n -tiles containing e in their boundary. We know that near each point of $\text{int}(e')$ the map f^k is an orientation-preserving homeomorphism. As in the proof of Lemma 5.2 one can use this fact to show that $f^k(X')$ and $f^k(Y')$ are distinct n -tiles. Since there are only two n -tiles containing e in their boundary, we conclude that one of the n -tiles $f^k(X')$ or $f^k(Y')$ is equal to X , say $f^k(X') = X$. Since $p \in e' \subset X'$, it follows that $p \in M$.

Hence $f^{-k}(X) = f^{-k}(\text{int}(X)) \cup f^{-k}(\partial X) \subset M$, and we conclude that $M = f^{-k}(X)$ as desired.

(iii) The 0-skeleton of \mathcal{D}^n is the set \mathbf{V}^n of all vertices of \mathcal{D}^n . By (ii) we know that $\mathbf{V}^n = f^{-n}(\mathbf{V}^0) = f^{-n}(\text{post}(f))$. Moreover,

$$f^{n+k}(\mathbf{V}^n) \subset f^{n+k}(f^{-n}(\text{post}(f))) \subset f^k(\text{post}(f)) \subset \text{post}(f) = \mathbf{V}^0,$$

and so $\mathbf{V}^n \subset \mathbf{V}^{n+k}$.

The 1-skeleton of \mathcal{D}^n is equal to the set consisting of all n -vertices and the union of all n -edges. As follows from (ii) this set is equal to the preimage of the 1-skeleton of \mathcal{D}^0 under the map f^n . Since the 1-skeleton of \mathcal{D}^0 is equal to \mathcal{C} , it follows that the 1-skeleton of \mathcal{D}^n is equal to $f^{-n}(\mathcal{C})$.

(iv) Note that $\deg(f^n) = \deg(f)^n$, and that $\#\mathbf{V}^0 = m$, $\#\mathbf{E}^0 = m$, and $\#\mathbf{X}^0 = 2$. The statements about \mathbf{V}^n , \mathbf{E}^n , and \mathbf{X}^n , then follow from the corresponding statement established in the last part of the proof of Lemma 5.4.

(v) This immediately follows from (iii) and Lemma 4.5.

(vi) If X is an n -tile, then $f^n|_X$ is a homeomorphism of X onto the 0-tile $f^n(X)$. The n -vertices contained in X are precisely the preimages of the 0-vertices contained in $f^n(X)$; hence X contains exactly $m = \#\text{post}(f)$ n -vertices, and hence also the same number of n -edges (Lemma 5.1 (iii)). So every n -tile is an m -gon. \square

Instead of an inequality for $\#\mathbf{V}^n$ as in (iv) one can easily give a precise formula for this number; namely, if we set $d = \deg(f)$, and $m = \#\text{post}(f)$, then $\#\mathbf{X}^n = 2d^n$ and $\mathbf{E}^n = md^n$. Moreover, by Euler's polyhedral formula we have

$$\#\mathbf{X}^n - \#\mathbf{E}^n + \#\mathbf{V}^n = 2,$$

and so

$$\#\mathbf{V}^n = (m - 2)d^n + 2.$$

By property (v) in the previous proposition we have

$$\text{mesh}(f, n, \mathcal{C}) = \max_{X \in \mathbf{X}^n} \text{diam}(X),$$

and so if \mathcal{C} is as in Definition 3.2, then f is expanding if

$$\lim_{n \rightarrow \infty} \max_{X \in \mathbf{X}^n} \text{diam}(X) = 0.$$

In other words, the Thurston map f is expanding if there exists a Jordan curve $\mathcal{C} \supset \text{post}(f)$ such that the diameter of the n -tiles for f and \mathcal{C} go to 0 uniformly with n . This fact was the motivation behind our definition of an expanding Thurston map.

It is often useful, in particular in graphical representations, to assign to each tile one of the two colors “black” and “white” represented by the symbols \mathbf{b} and \mathbf{w} , respectively. To formulate this, we denote by \mathbf{X}^∞ the disjoint union of the sets \mathbf{X}^n , $n \in \mathbb{N}_0$ (for given f and \mathcal{C}). More informally, \mathbf{X}^∞ is the set of all tiles. Note that in general, a set can be a tile for different levels n , so the same tile may be represented by multiple copies in \mathbf{X}^∞ distinguished by their levels n .

Lemma 6.2 (Colors of tiles). *There exists a map $L: \mathbf{X}^\infty \rightarrow \{\mathbf{b}, \mathbf{w}\}$ with the following properties:*

- (i) $L(X_{\mathbf{b}}^0) = \mathbf{b}$ and $L(X_{\mathbf{w}}^0) = \mathbf{w}$.
- (ii) If $n, k \in \mathbb{N}_0$, $X^{n+k} \in \mathbf{X}^{n+k}$, and $X^n = f^k(X^{n+k}) \in \mathbf{X}^n$, then $L(X^n) = L(X^{n+k})$.
- (iii) If $n \in \mathbb{N}_0$, and X^n and Y^n are two distinct n -tiles that have an n -edge in common, then $L(X^n) \neq L(Y^n)$.

Moreover, L is uniquely determined by properties (i) and (ii).

So with the normalization (i) one can uniquely assign colors “black” or “white” to the tiles so that all iterates of f are color-preserving as in (ii). By (iii) colors of distinct n -tiles are different if they share an n -edge.

Our notion of colorings of tiles is related to the more general concept of a *labeling* of cells in a cell decomposition (see Section 12, in particular Lemma 12.6).

Proof. To define L we assign colors to the two 0-tiles $X_{\mathbf{b}}^0$ and $X_{\mathbf{w}}^0$ as in (i). If Z^n is an n -tile for some arbitrary level $n \geq 0$, then $f^n(Z^n)$ is a 0-tile (Proposition 6.1 (i)), and so it already has a color assigned. We set $L(Z^n) := L(f^n(Z^n))$.

This defines a map $L: \mathbf{X}^\infty \rightarrow \{\mathbf{b}, \mathbf{w}\}$. By definition, L has property (i). To show (ii), assume that $n, k \in \mathbb{N}_0$ and $X^{n+k} \in \mathbf{X}^{n+k}$. Then by Proposition 6.1 (i), we have $X^n := f^k(X^{n+k}) \in \mathbf{X}^n$, and $f^{n+k}(X^{n+k}), f^n(X^n) \in \mathbf{X}^0$. So by definition of L we have

$$L(X^n) = L(f^n(X^n)) = L(f^n(f^k(X^{n+k}))) = L(f^{n+k}(X^{n+k})) = L(X^{n+k})$$

as desired.

Let X^n and Y^n be as in (iii). Then again by Proposition 6.1 (i), we have $f^n(X^n), f^n(Y^n) \in \mathbf{X}^0$. Moreover, by the same argument as in Case 2 of the proof of Lemma 5.2 the 0-tiles $f^n(X^n)$ and $f^n(Y^n)$ are distinct. So one of them is equal to $X_{\mathbf{b}}^0$, while the other one is equal to $X_{\mathbf{w}}^0$. In particular, $L(f^n(X^n)) \neq L(f^n(Y^n))$, and so by definition of L we have

$$L(X^n) = L(f^n(X^n)) \neq L(f^n(Y^n)) = L(Y^n)$$

as desired. It follows that L has the properties (i)–(iii).

It is clear that L is uniquely determined by (i) and (ii). \square

By using the cell decompositions \mathcal{D}^n one can easily classify all Thurston maps with two postcritical points up to Thurston equivalence.

Proposition 6.3. *Let $f: S^2 \rightarrow S^2$ be a Thurston map with $\#\text{post}(f) = 2$. Then f is Thurston equivalent to a map of the form $z \mapsto z^k$ on $\widehat{\mathbb{C}}$, where $k \in \mathbb{Z} \setminus \{-1, 0, 1\}$.*

Proof. Using some auxiliary conjugations if necessary, we may assume that $S^2 = \widehat{\mathbb{C}}$ and $\text{post}(f) = \{0, \infty\}$. We pick $\mathcal{C} = \widehat{\mathbb{R}} \supset \text{post}(f)$, and consider the cell decompositions $\mathcal{D}^n(f, \mathcal{C})$ of $\widehat{\mathbb{C}}$. Since $\text{post}(f) \subset f^{-n}(\text{post}(f))$, the points 0 and ∞ are n -vertices for each $n \in \mathbb{N}_0$. Consider the 1-tiles X_1, \dots, X_d , and the 1-edges e_1, \dots, e_d of the cycle of 0, considered as a 1-vertex, where d is the length of the cycle. If the indexing of these tiles and edges is as in Lemma 5.1 (v), then

$$(6.1) \quad \partial X_j = e_{j-1} \cup e_j$$

for $j = 1, \dots, d$ (where $e_0 = e_d$). This is true, because $\#\text{post}(f) = 2$ and so every n -tile is a 2-gon. It shows that apart from 0 the edges e_1, \dots, e_d have one other 1-vertex $p \neq 0$ in common.

It follows from Lemma 5.1 (iv) that the set

$$U = \bigcup_{j=1}^d \text{int}(X_j) \cup \bigcup_{i=j}^d \text{int}(e_i)$$

is open, and (6.1) implies that $\partial U = \{0, p\}$. Hence 0 and p are isolated boundary points of U , and so the set $\overline{U} = U \cup \{0, p\}$ is open and closed. We conclude that $\overline{U} = S^2$. Since ∞ is a 1-vertex, and the set U does not contain any 1-vertex, we must have $p = \infty$.

Since $f(\text{post}(f)) \subset \text{post}(f)$, we have $f(0) = 0$ or $f(0) = \infty$. Assume that $f(0) = 0$. Since f is injective on 1-tiles and $f(\infty) \in \{0, \infty\}$, we have $f(\infty) = \infty$. By an argument similar to (and simpler than) the one in the proof of Lemma 5.2 we will show that one can find a homeomorphism $\varphi: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ that fixes 0 and ∞ and satisfies $f(\varphi(z)) = z^k$ for all $z \in \widehat{\mathbb{C}}$, where $k = d/2$. Note that d must be even, because f maps the 1-tiles X_1, \dots, X_d alternately to the two 0-tiles, i.e., the upper and the lower half-planes in $\widehat{\mathbb{C}}$. By reindexing if necessary, we may assume that X_1, X_3, \dots are mapped to the upper and X_2, X_4, \dots to the lower half-plane. Define sectors in $\widehat{\mathbb{C}}$ by

$$\Sigma_j = \{re^{2\pi it} : (j-1)/d \leq t \leq j/d, r \geq 0\} \cup \{\infty\}$$

for $j = 1, \dots, d$. If $z \in \widehat{\mathbb{C}}$ is arbitrary, then $z \in \Sigma_j$ for some $j \in \{1, \dots, d\}$. Then z^k lies in the upper or lower half-plane depending on whether j is even or odd; hence $(f|X_j)^{-1}(z^k)$ is defined. We put $\varphi(z) := (f|X_j)^{-1}(z^k)$. Then φ is a well-defined homeomorphism on $\widehat{\mathbb{C}}$ that fixes 0 and ∞ and satisfies $f(\varphi(z)) = z^k$ for all $z \in \widehat{\mathbb{C}}$. This last identity implies that φ is orientation-preserving, because the maps f and $z \mapsto z^k$ are a local orientation-preserving homeomorphisms away from their critical points.

If $f(0) = \infty$, then we apply the above argument to the map $z \mapsto 1/f(z)$. We conclude that in any case there exists an orientation-preserving homeomorphism φ on $\widehat{\mathbb{C}}$ that fixes 0 and ∞ and satisfies $f(\varphi(z)) = z^k$ for all $z \in \widehat{\mathbb{C}}$, where k is a non-zero integer. Actually, $k \notin \{-1, 1\}$, because otherwise f would be a homeomorphism. Since every orientation-preserving homeomorphism φ on $\widehat{\mathbb{C}}$ fixing 0 and ∞ is isotopic to $\text{id}_{\widehat{\mathbb{C}}}$ rel. $\{0, \infty\}$ (see Lemma 10.11), it follows that f is Thurston equivalent to the map $z \mapsto z^k$, where $k \in \mathbb{Z} \setminus \{-1, 0, 1\}$. \square

Corollary 6.4. *If $f: S^2 \rightarrow S^2$ is an expanding Thurston map, then $\# \text{post}(f) \geq 3$.*

Proof. We know that $\# \text{post}(f) \geq 2$ (see Remark 5.5). The same reasoning as in the proof of Proposition 6.3 shows that if $f: S^2 \rightarrow S^2$ is a Thurston map with $\# \text{post}(f) = 2$, then each n -tile contains the set $\text{post}(f)$. In particular, n -tiles have a diameter uniformly bounded away from 0, and so the map cannot be expanding. Therefore, if f is expanding, then $\# \text{post}(f) \geq 3$. \square

Due to the last corollary, in the following we can restrict ourselves to the case of Thurston maps f with $\# \text{post}(f) \geq 3$.

7. FLOWERS

In this section $f: S^2 \rightarrow S^2$ is a Thurston map with $\# \text{post}(f) \geq 3$. We fix a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$, and consider the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ and use the terminology and notation of the previous section.

Definition 7.1 (n -Flowers). Let $n \in \mathbb{N}_0$, and $p \in S^2$ be an n -vertex. Then the n -flower of p is defined as

$$W^n(p) := \bigcup \{ \text{int}(c) : c \in \mathcal{D}^n, p \in c \}.$$

So the n -flower $W^n(p)$ of the n -vertex p is the union of the interiors of all cells in cycle of p in \mathcal{D}^n (see Figure 2 as well as Lemma 5.1 (v) and the discussion after this lemma).

The main reason why we introduced flowers is the following. Consider a simply connected domain $U \subset S^2$ not containing a postcritical point of f and branches g_n of f^{-n} defined on U . Then it may happen that the number of n -tiles intersecting $g_n(U)$ is unbounded as $n \rightarrow \infty$, even if the diameter of U is small. For example, this happens when f has a periodic critical point p (see Section 17), and U spirals around one of the points in the cycle generated by p . However if $\text{diam}(U)$ is sufficiently small, then $g_n(U)$ is always contained in one n -flower as we shall see. We first prove some basic properties of flowers.

Lemma 7.2. *Let $n \in \mathbb{N}_0$, and $p \in S^2$ be an n -vertex. As in Lemma 5.1 let e_1, \dots, e_d be the n -edges and X_1, \dots, X_d be the n -tiles of the cycle of p , where $d \in \mathbb{N}$, $d \geq 2$, is the length of the cycle.*

- (i) *Then $d = 2 \deg_{f^n}(p)$, the set $W^n(p)$ is an open and simply connected neighborhood of p that contains no other n -vertex, and we have*

$$(7.1) \quad W^n(p) = \{p\} \cup \bigcup_{i=1}^d \text{int}(X_i) \cup \bigcup_{i=1}^d \text{int}(e_i) = S^2 \setminus \bigcup \{c \in \mathcal{D}^n : c \in \mathcal{D}^n, p \notin c\}.$$

- (ii) *We have $\overline{W^n(p)} = X_1 \cup \dots \cup X_d$, and the set $\partial W^n(p)$ is the union of all n -edges e with $p \notin e$ and $e \subset \partial X_i$ for some $i \in \{1, \dots, d\}$.*
- (iii) *If c is an arbitrary n -cell, then either $p \in c$ and $c \subset \overline{W^n(p)}$, or $c \subset S^2 \setminus W^n(p)$.*

Proof. (i) By Remark 5.3 the length d of the cycle of the vertex p (in the cell decomposition \mathcal{D}^n) is a multiple $d = kd'$ of the length d' of the cycle of the image point $q = f^n(p)$ (in the cell decomposition \mathcal{D}^0), where k is the degree of f^n at p . Since $d' = 2$, we have $d = 2 \deg_{f^n}(p)$ as claimed.

The first equality in (7.1) follows from Lemma 5.1 (v). Based on this, the argument in Case 3 of the proof of Lemma 5.2 shows that the set $W^n(p)$ is homeomorphic to \mathbb{D} . Hence $W^n(p)$ is open and simply connected, and it follows from the first equality in (7.1) that $W^n(p)$ contains no other n -vertex than p .

Let $M = S^2 \setminus \bigcup \{c \in \mathcal{D}^n : c \in \mathcal{D}^n, p \notin c\}$. If $x \in W^n(p)$, then x is an interior point in one of the cells τ forming the cycle of p . So if c is any n -cell with $x \in c$, then $\tau \subset c$ by Lemma 4.3 (ii). This implies $p \in c$, and so $x \in M$ by definition of M . Hence $W^n(p) \subset M$.

Conversely, if $x \in M$, let τ be an n -cell of smallest dimension that contains x . Obviously, $x \in \text{int}(\tau)$. On the other hand, the definition

of M implies that $p \in \tau$. Hence τ is a cell in the cycle of p , and so $x \in W^n(p)$. We conclude $M \subset W^n(p)$, and so $M = W^n(p)$ as desired.

(ii) Equation (7.1) implies $\overline{W^n(p)} = X_1 \cup \dots \cup X_n$.

Every point $x \in \partial W^n(p)$ is contained in one of the sets ∂X_i . Since $W^n(p)$ is open, the point x is not contained in $\{p\} \cup \text{int}(e_{i-1}) \cup \text{int}(e_i) \subset W^n(p)$ and hence in one of the n -edges e in the boundary of X_i distinct from e_{i-1} and e_i ; note that there exists such an edge, because each n -tile is an m -gon, where $m = \# \text{post}(f) \geq 3$, and so contains more than two n -edges in its boundary. Then $p \notin e$, and x is contained in an n -edge with the desired properties.

Conversely, if e is an n -edge with $p \notin e$ and $e \subset \partial X_i$, then $e \subset S^2 \setminus W^n(p)$ by (7.1), and $e \subset X_i \subset \overline{W^n(p)}$. Hence $e \subset \partial W^n(p)$.

(iii) This follows from (i) and (7.1). \square

Note that if we color tiles as in Lemma 6.2, then the colors of the tiles X_1, \dots, X_d associated to an n -flower as in the previous lemma will alternate.

Lemma 7.3. *Let $k, n \in \mathbb{N}_0$.*

- (i) *If $p \in S^2$ is an $(n+k)$ -vertex, then $f^k(W^{n+k}(p)) = W^n(q)$.*
- (ii) *If $q \in S^2$ is an n -vertex, then the connected components of $f^{-k}(W^n(q))$ are the $(n+k)$ -flowers $W^{n+k}(p)$, $p \in f^{-k}(q)$.*
- (iii) *A connected set $K \subset S^2$ is contained in an $(n+k)$ -flower if and only if $f^k(K)$ is contained in an n -flower.*

Proof. (i) It is clear that $q = f^k(p)$ is an n -vertex. Let $e'_1, \dots, e'_{d'}$ be the $(n+k)$ -edges and $X'_1, \dots, X'_{d'}$ be the $(n+k)$ -tiles in the cycle of p , and define $e_i = f^k(e'_i)$ and $X_i = f^k(X'_i)$ for $i = 1, \dots, d'$. Then from Remark 5.3 it follows that $e_1, \dots, e_{d'}$ are the n -edges and $X_1, \dots, X_{d'}$ are the n -tiles in the cycle of q . Here we may have possible repetitions of edges and tiles. Since the map f^k is cellular for $(\mathcal{D}^{n+k}, \mathcal{D}^n)$, we have $f^k(\text{int}(e'_i)) = \text{int}(e_i)$ and $f^k(\text{int}(X'_i)) = \text{int}(X_i)$ for all $i = 1, \dots, d'$. Using this and (7.1) the statement follows.

(ii) If $p \in f^{-1}(q)$, then p is an $(n+k)$ -vertex. By (i) the $(n+k)$ -flower $W^{n+k}(p)$ is an open and connected subset of $f^{-k}(W^n(q))$. Suppose that $x \in \partial W^{n+k}(p)$. Then by Lemma 7.2 (ii) there exists an $(n+k)$ -tile X' , and an $(n+k)$ -edge e' with $p \in X'$, $p \notin e'$, and $x \in e' \subset \partial X'$. Then $X = f^k(X')$ is an n -tile, $e = f^k(e')$ is an n -edge, and $q \in X$, $f(x) \in e \subset \partial X$. Since $f^k|_{X'}$ is a homeomorphism of X' onto X , we also have $q \notin X$. Lemma 7.2 (ii) implies that $f^k(x) \in \partial W^n(q)$, and so $f^k(x) \notin W^n(q)$, because flowers are open sets.

We conclude that $x \in S^2 \setminus f^{-k}(W^n(q))$, and so $\partial W^{n+k}(p) \subset S^2 \setminus f^{-k}(W^n(q))$. It now follows from Lemma 4.4 that $W^{n+k}(p)$ is a connected component of $f^{-k}(W^n(q))$.

Conversely, suppose that U is a connected component of $f^{-k}(W^n(q))$. Then U is an open set and so it meets the interior $\text{int}(X')$ of some $(n+k)$ -tile X' . Then $X = f^k(X')$ is an n -tile that meets $W^n(q)$. Hence $q \in X$, and so there exists an $(n+k)$ -vertex $p \in X'$ with $f^k(p) = q$.

Then by the first part of the proof, the set $W^{n+k}(p)$ is a connected component of $f^{-k}(W^n(q))$. Since $W^{n+k}(p)$ contains the set $\text{int}(X')$ and so meets U , we must have $W^{n+k}(p) = U$.

(iii) Suppose K is contained in the $(n+k)$ -flower $W^{n+k}(p)$. Then by (i) the set $f^k(W^{n+k}(p)) = W^n(f^k(p))$ is an n -flower and it contains $f^k(K)$.

Conversely, if $f^k(K)$ is contained in the n -flower $W^n(q)$, then K is a connected set in $f^{-k}(W^n(q))$. Hence K lies in a connected component of $f^{-k}(W^n(q))$, and hence in an $(n+k)$ -flower by (ii). \square

Similarly, as we defined an n -flower for an n -vertex, one can also define an *edge flower* for an n -edge. These sets provide “canonical” neighborhoods for n -vertices and n -edges defined in terms of n -cells.

Definition 7.4 (Edge flowers). Let $n \in \mathbb{N}_0$, and e be an n -edge. Then the *n -edge flower* of e is defined as

$$W^n(e) := \bigcup \{\text{int}(c) : c \in \mathcal{D}^n, c \cap e \neq \emptyset\}.$$

We list some properties of edge flowers. They correspond to similar properties of n -flowers as in Lemma 7.2. Note that in contrast to an n -flower, an edge flower will not be simply connected in general.

Lemma 7.5. *Let e be an n -edge whose boundary ∂e consists of the n -vertices u and v .*

(i) *Then $W^n(e)$ is an open set containing e , and*

$$(7.2) \quad W^n(e) = W^n(u) \cup W^n(v) = S^2 \setminus \bigcup \{c : c \in \mathcal{D}^n, c \cap e = \emptyset\}.$$

(ii) *We have $\overline{W^n(e)} = \bigcup \{X \in \mathbf{X}^n : X \cap e \neq \emptyset\}$ and*

$$\partial W^n(e) = \bigcup \{c \in \mathcal{D}^n : c \cap e = \emptyset \text{ and}$$

there exists $X \in \mathbf{X}^n$ with $X \cap e \neq \emptyset$ and $c \subset \partial X$ \},

where each n -cell c in the last union either consists of one n -vertex or is an n -edge.

(iii) *If c is an arbitrary n -cell, then either $c \cap e \neq \emptyset$ and $c \subset \overline{W^n(e)}$, or $c \subset S^2 \setminus W^n(e)$.*

Proof. (i) It follows from Lemma 4.3 (i) that an n -cell c meets e if and only if it contains one of the endpoints u and v of e . Hence $W^n(e) = W^n(u) \cup W^n(v)$ by the definition of flowers. By Lemma 7.2 (i) this implies that $W^n(e)$ is open, and, since e is an edge in the cycles of u and v , we also have

$$e = \{u\} \cup \text{int}(e) \cup \{v\} \subset W^n(u) \cup W^n(v) = W^n(e).$$

Let $M = S^2 \setminus \bigcup\{c : c \in \mathcal{D}^n, c \cap e = \emptyset\}$. If an n -cell c does not meet e , then it contains neither u nor v . Hence by (7.1) we have $S^2 \setminus M \subset (S^2 \setminus W^n(u)) \cap (S^2 \setminus W^n(v)) = S^2 \setminus W^n(e)$, and so $W^n(e) \subset M$.

Conversely, let $x \in M$ be arbitrary, and c be the unique n -cell c such that $x \in \text{int}(c)$. Then $c \cap e \neq \emptyset$ and so $u \in c$ or $v \in c$. It follows that $x \in W^n(u) \cup W^n(v) = W^n(e)$. We conclude that $M \subset W^n(e)$, and so $M = W^n(e)$ as claimed.

(ii) By Lemma 4.3 (i) an n -tile X meets e if and only if X contains u or v . Hence by (i) and Lemma 7.10 (ii) we have

$$\overline{W^n(e)} = \overline{W^n(u)} \cup \overline{W^n(v)} = \bigcup\{X \in \mathbf{X}^n : X \cap e \neq \emptyset\}$$

as desired.

For the second claim suppose that c is an n -cell and X an n -tile with $c \cap e = \emptyset$, $X \cap e \neq \emptyset$, and $c \subset \partial X$. Then $c \subset S^2 \setminus W^n(e)$ and c must be an n -edge or consist of an n -vertex. Moreover, $c \subset X \subset \overline{W^n(e)}$. It follows that $c \subset \partial W^n(e)$.

Conversely, let x be a point in $\partial W^n(e)$. Then by (i) the point x is also a boundary point of $W^n(u)$ or $W^n(v)$, say $x \in \partial W^n(u)$.

By Lemma 7.2 (ii) there exists an n -edge e' and an n -tile X with $x \in e'$, $u \in X$, $u \notin e'$ and $e' \subset \partial X$. If x is an n -vertex, we let $c = \{x\}$. Then c is an n -cell and we have $c \cap e = \emptyset$, because $W^n(e)$ is an open neighborhood of e and c lies in $\partial W^n(e) \subset S^2 \setminus W^n(e)$. Moreover, $X \cap e \neq \emptyset$ and $c \subset e' \subset \partial X$. So c is an n -cell with the desired properties containing x .

If x is not a vertex we put $c = e'$. Again if $c \cap e = e' \cap e = \emptyset$, then c is an n -cell with the desired properties containing x .

The other case, where $e' \cap e \neq \emptyset$, leads to a contradiction. Indeed, then we have $v \in e'$. Moreover, since x is not a vertex, it follows that $x \in \text{int}(e')$; but then $x \in \text{int}(e') \subset W^n(v) \subset W^n(e)$ which is impossible, because $x \in \partial W^n(e) \subset S^2 \setminus W^n(e)$.

(iii) If c is an n -cell and $c \cap e = \emptyset$, then $c \subset S^2 \setminus W^n(e)$. If $c \cap e \neq \emptyset$, then c contains u or v , and so $c \subset \overline{W^n(u)} \cup \overline{W^n(v)} = \overline{W^n(e)}$. \square

We fix a base metric on S^2 that induces the given topology. We will define a constant $\delta_0 > 0$ such that any connected set of diameter $< \delta_0$

(with respect to the base metric) is contained in a single 0-flower. There is a slight difference for the cases $\# \text{post}(f) = 3$ and $\# \text{post}(f) \geq 4$. In order to treat these two cases simultaneously, the following definition is useful.

Definition 7.6 (Joining opposite sides). A set $K \subset S^2$ joins opposite sides of \mathcal{C} if $\# \text{post}(f) \geq 4$ and K meets two disjoint 0-edges, or if $\# \text{post}(f) = 3$ and K meets all three 0-edges.

We then define

$$(7.3) \quad \delta_0 = \delta_0(f, \mathcal{C}) = \inf \{ \text{diam}(K) : K \subset S^2 \text{ is a set joining opposite sides of } \mathcal{C} \}.$$

Then $\delta_0 > 0$. For if $\# \text{post}(f) = 4$, then δ_0 is bounded below by the positive number

$$\min \{ \text{dist}(e, e') : e \text{ and } e' \text{ are disjoint 0-edges} \}.$$

If $\# \text{post}(f) = 3$ and we had $\delta_0 = 0$, then it would follow from a simple limiting argument that the three 0-edges had a common point. This is absurd.

Lemma 7.7. *A connected set $K \subset S^2$ joins opposite sides of \mathcal{C} if and only if K is not contained in a single 0-flower (of a 0-vertex).*

Proof. If K is contained in a 0-flower $W^0(p)$, where $p \in \mathcal{C}$ is a 0-vertex, then K meets at most two 0-edges, namely the ones that have the common endpoint p . So K does not join opposite sides of \mathcal{C} .

Conversely, suppose K does not join opposite sides of \mathcal{C} . We have to show that K is contained in some 0-flower. Note that K cannot meet three distinct 0-edges.

If K does not meet any 0-edge, then K is contained in every 0-flower. If K meets only one 0-edge e , then K is contained in the 0-flowers $W^0(u)$ and $W^0(v)$, where u and v are the endpoints of e .

If K meets two edges, then these edges share a common endpoint $v \in \mathbf{V}^0 = \text{post}(f)$. This is always true if $\# \text{post}(f) = 3$ and follows from the fact that K does not join opposite sides of \mathcal{C} if $\# \text{post}(f) \geq 4$. Moreover, K cannot meet a third 0-edge which implies that $K \subset W^n(v)$. \square

By the previous lemma every connected set $K \subset S^2$ with $\text{diam}(K) < \delta_0$ is contained in a 0-flower.

Lemma 7.8. *Let $n \in \mathbb{N}_0$, and $\delta_0 > 0$ be as in (7.3).*

- (i) If $K \subset S^2$ is a connected set with $\text{diam}(K) < \delta_0$, then each component of $f^{-n}(K)$ is contained in some n -flower.
- (ii) If $\gamma: [0, 1] \rightarrow S^2$ is a path such that $\text{diam}(\gamma) < \delta_0$, then each lift $\tilde{\gamma}$ of γ by f^n has an image that is contained in some n -flower.

Here by definition a *lift* of γ by f^n is any path $\tilde{\gamma}: [0, 1] \rightarrow S^2$ with $\gamma = f^n \circ \tilde{\gamma}$.

Proof. (i) The set K is contained in some 0-flower $W^0(p)$, $p \in \mathbf{V}^0$, by Lemma 7.7 and the definition of δ_0 . So if K' is a component of $f^{-n}(K)$, then K' is contained in a component of $f^{-n}(W^0(p))$, and hence in an n -flower by Lemma 7.3 (ii).

(ii) The reasoning is exactly the same as in (i). The image of γ is contained in some 0-flower; by Lemma 7.3 (ii) this implies that the image of $\tilde{\gamma}$ is contained in an n -flower. \square

We will often have to estimate how many tiles are needed to connect certain points. If we have a condition that is formulated “at the top level”, i.e., for connecting points in \mathcal{C} , then the map f^n can be used to translate this to n -tiles.

Lemma 7.9. *Let $n \in \mathbb{N}_0$, and $K \subset S^2$ be a connected set. If there exist two disjoint n -cells σ and τ with $K \cap \sigma \neq \emptyset$ and $K \cap \tau \neq \emptyset$, then $f^n(K)$ joins opposite sides of \mathcal{C} .*

Proof. It suffices to show that K is not contained in any n -flower, because then $f^n(K)$ is not contained in any 0-flower (Lemma 7.3 (iii)) and so $f^n(K)$ joins opposite sides of \mathcal{C} (Lemma 7.7). We consider several cases.

Case 1. One of the cells is an n -vertex, say $\sigma = \{v\}$, where $v \in \mathbf{V}^n$. Then $v \in K$, so the only n -flower that K could possibly be contained in is $W^n(v)$, because no other n -flower contains the n -vertex v . But since σ and τ are disjoint, we have $v \notin \tau$, and so $\tau \subset S^2 \setminus W^n(v)$. Hence $K \cap (S^2 \setminus W^n(v)) \neq \emptyset$, and so $W^n(v)$ does not contain K .

Case 2. Suppose one of the cells is an n -edge, say $\sigma = e \in \mathbf{E}^n$. Then e has two endpoints $u, v \in \mathbf{V}^n$. The only n -flowers that meet e are $W^n(u)$ and $W^n(v)$; so these n -flowers are the only ones that could possibly contain K . But the set $W^n(e) = W^n(u) \cup W^n(v)$ does not contain K , because K meets the set τ which lies in the complement of $W^n(e)$.

Case 3. One of the cells in an n -tile, say $\sigma \in \mathbf{X}^n$. Then K meets ∂X . Since ∂X consists of n -edges, the set K meets an n -edge disjoint from τ . So we are reduced to Case 2. \square

For $n \in \mathbb{N}_0$ we denote by D_n the minimal number of n -tiles required to form a connected set joining opposite sides of \mathcal{C} ; more precisely,

$$(7.4) \quad D_n = \min \{N \in \mathbb{N} : \text{there exist } X_1, \dots, X_N \in \mathbf{X}^n \text{ such that}$$

$$K = \bigcup_{j=1}^N X_j \text{ is connected and joins opposite sides of } \mathcal{C}\}.$$

Of course, D_n depends on f and the choice of \mathcal{C} . If we want to emphasize this dependence, we write $D_n = D_n(f, \mathcal{C})$.

From Lemma 7.9 we can immediately derive the following consequence.

Lemma 7.10. *Let $n, k \in \mathbb{N}_0$. Every set of $(n+k)$ -tiles whose union is connected and meets two disjoint n -cells contains at least D_k elements.*

Proof. Suppose K is a union of $(n+k)$ -tiles with the stated properties. Then the images of these tiles under f^n are k -tiles and $f^n(K)$ joins opposite sides of \mathcal{C} by Lemma 7.9. Hence there exist at least D_k distinct k -tiles in the union forming $f^n(K)$ and hence at least D_k distinct $(n+k)$ -tiles in K . \square

Lemma 7.11. *There exists $M \in \mathbb{N}$ with the following property:*

- (i) *Each n -tile, $n \in \mathbb{N}$, can be covered by M $(n-1)$ -flowers.*
- (ii) *Each n -tile, $n \in \mathbb{N}_0$, can be covered by M $(n+1)$ -flowers.*

For easier formulation of this lemma and the subsequent proof, we assume for simplicity that a cover by *at most* M element contains precisely M elements. This can always be achieved by repetition of elements in the cover.

Proof. (i) Let $\delta_0 > 0$ be as in (7.3). Then there exists $M \in \mathbb{N}$ such that each of the finitely many 1-tiles X is a union of M connected sets $U \subset X$ with $\text{diam}(U) < \delta_0$. If Y is an arbitrary n -tile, $n \geq 1$, then $Z = f^{n-1}(Y)$ is a 1-tile and $f^{n-1}|_Y$ a homeomorphism of Y onto Z . Hence Y is a union of M sets of the form $(f^{n-1}|_Y)^{-1}(U)$, where $U \subset Z$ is connected and $\text{diam}(U) < \delta_0$. Each set $(f^{n-1}|_Y)^{-1}(U)$ is connected and so by Lemma 7.8 (i) it lies in an $(n-1)$ -flower. Hence Y can be covered by M $(n-1)$ -flowers.

(ii) There exists $M \in \mathbb{N}$ such that each of the two 0-tiles X can be covered by M connected sets $U \subset X$ with $\text{diam}(f(U)) < \delta_0$. If Y is an arbitrary n -tile, then $Z = f^n(Y)$ is a 0-tile. By the same reasoning as above, the set Y is a union of M sets of the form $(f^n|_Y)^{-1}(U)$, where $U \subset Z$ is connected and $\text{diam}(f(U)) < \delta_0$.

Then $U' = (f^n|Y)^{-1}(U)$ is connected, and $f^{n+1}(U') = f(U)$ which implies $\text{diam}(f^{n+1}(U')) < \delta_0$. Hence by Lemma 7.8 (i) the set U' is contained in some $(n+1)$ -flower. Since M of the sets U' cover Y , it follows that each n -tile can be covered by M $(n+1)$ -flowers. \square

Lemma 7.12. *Let \mathcal{C} and $\tilde{\mathcal{C}}$ be two Jordan curves in S^2 that both contain $\text{post}(f)$. Then there exists a number M such that each n -tile for $(f, \tilde{\mathcal{C}})$ is covered by M n -flowers for (f, \mathcal{C}) .*

Proof. The proof is very similar to the proof of Lemma 7.11.

Let $\delta_0 = \delta_0(f, \mathcal{C}) > 0$ be the number as defined in (7.3). There exists a number M such that each of the two 0-tiles X for $(f, \tilde{\mathcal{C}})$ is a union of M connected sets $U \subset X$ with $\text{diam}(U) < \delta_0$. If Y is an arbitrary n -tile for $(f, \tilde{\mathcal{C}})$, then $Z = f^n(Y)$ is a 0-tile for $(f, \tilde{\mathcal{C}})$ and $f^n|Y$ is a homeomorphism of Y onto Z . Hence Y is a union of M sets of the form $(f^n|Y)^{-1}(U)$, where $U \subset Z$ is connected and $\text{diam}(U) < \delta_0$. Each set $(f^n|Y)^{-1}(U)$ is connected and so by Lemma 7.8 (i) it lies in an n -flower for (f, \mathcal{C}) . Hence Y can be covered by M such n -flowers. \square

8. EXPANSION AND VISUAL METRICS

Let $f: S^2 \rightarrow S^2$ be a Thurston map. Throughout this section we will assume that f is expanding. We will show that S^2 then carries a natural class of metrics that allows us to estimate the distance of points in terms of combinatorial data derived from the tiles in the cell decompositions defined in Section 6.

We fix a base metric on S^2 that induces the standard topology. The purpose of this is to be able to formulate some essentially topological properties (such as expansion of the map f) in more convenient metric terms. Notation for metric terms will refer to this base metric unless otherwise indicated.

As we have seen in Section 6, the property of expansion can equivalently be stated as

$$(8.1) \quad \max_{X \in \mathbf{X}^n} \text{diam}(X) \rightarrow 0 \text{ as } n \rightarrow \infty,$$

where the tiles are defined with respect to a Jordan curve $\mathcal{C} \supset \text{post}(f)$ as in Definition 3.2. First we want to convince ourselves that this is independent of the choice of the curve \mathcal{C} .

Lemma 8.1. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map. Then*

$$(8.2) \quad \lim_{n \rightarrow \infty} \text{mesh}(f, n, \tilde{\mathcal{C}}) = 0$$

for every Jordan curve $\tilde{\mathcal{C}} \subset S^2$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$.

Proof. Since f is expanding, there exists a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$ such that (8.2) holds. Then

$$\max_{X \in \mathbf{X}^n} \text{diam}(X) = \text{mesh}(f, n, \mathcal{C}) \rightarrow 0$$

as $n \rightarrow \infty$, where \mathbf{X}^n is the set of n -tiles for (f, \mathcal{C}) . Lemma 7.2 (ii) implies that

$$(8.3) \quad \text{diam}(W^n(v)) \leq 2 \max_{X \in \mathbf{X}^n} \text{diam}(X)$$

for each n -flower for (f, \mathcal{C}) .

Let $\tilde{\mathcal{C}} \subset S^2$ be another Jordan curve with $\text{post}(f) \subset \tilde{\mathcal{C}}$, and the number M be as in Lemma 7.12. Then each n -tile for $(f, \tilde{\mathcal{C}})$ can be covered by M n -flowers for (f, \mathcal{C}) . If a connected set is covered by a finite union of connected sets, then its diameter is bounded by the sum of the diameters of the sets in the union. Combining this with (8.3) and denoting the set of n -vertices for (f, \mathcal{C}) by \mathbf{V}^n , we conclude that

$$\begin{aligned} \text{mesh}(f, n, \tilde{\mathcal{C}}) &= \max\{\text{diam}(\tilde{X}) : \tilde{X} \text{ is an } n\text{-tile for } (f, \tilde{\mathcal{C}})\} \\ &\leq M \max_{p \in \mathbf{V}^n} \text{diam}(W^n(p)) \\ &\leq 2M \max_{X \in \mathbf{X}^n} \text{diam}(X) \\ &= 2M \text{mesh}(f, n, \mathcal{C}). \end{aligned}$$

Hence $\text{mesh}(f, n, \tilde{\mathcal{C}}) \rightarrow 0$ as $n \rightarrow \infty$ as desired. \square

Our definition of expansion is somewhat *ad hoc*, but it has the advantage that it relates to the geometry of tiles. An equivalent, and maybe more conceptual description can be given in terms of the behavior of open covers of S^2 under pull-backs by the iterates of the map. We start with some definitions. Let \mathcal{U} be an open cover of S^2 . We define $\text{mesh}(\mathcal{U})$ to be the supremum of all diameters of connected components of sets in \mathcal{U} . If $g: S^2 \rightarrow S^2$ is a continuous map, then the *pull-back* of \mathcal{U} by g is defined as

$$g^{-1}(\mathcal{U}) = \{V : V \text{ connected component of } g^{-1}(U), \text{ where } U \in \mathcal{U}\}.$$

Obviously, $g^{-1}(\mathcal{U})$ is also an open cover of S^2 . Similarly we denote by $g^{-n}(\mathcal{U})$ the pull-back of \mathcal{U} by g^n .

Proposition 8.2. *Let $f: S^2 \rightarrow S^2$ be a Thurston map. Then the following conditions are equivalent:*

- (i) *The map f is expanding.*

- (ii) *There exists $\delta_0 > 0$ with the following property: if \mathcal{U} is a cover of S^2 by open and connected sets that satisfies $\text{mesh}(\mathcal{U}) < \delta_0$, then*

$$\lim_{n \rightarrow \infty} \text{mesh}(f^{-n}(\mathcal{U})) = 0.$$

- (iii) *There exists an open cover \mathcal{U} of S^2 with*

$$\lim_{n \rightarrow \infty} \text{mesh}(f^{-n}(\mathcal{U})) = 0.$$

- (iv) *There exists an open cover \mathcal{U} of S^2 with the following property: for every open cover \mathcal{V} of S^2 there exists $N \in \mathbb{N}$ such that $f^{-n}(\mathcal{U})$ is finer than \mathcal{V} for every $n \in \mathbb{N}$ with $n > N$; i.e., for every set $U' \in f^{-n}(\mathcal{U})$ there exists a set $V \in \mathcal{V}$ such that $U' \subset V$.*

Conditions (iii) is the notion of expansion as defined by Haïssinsky-Pilgrim (see [HP09, Sect. 2.2]). Thus our notion of expansion agrees with the one in [HP09]. Condition (iv) is essentially a reformulation of (iii) in purely topological terms without reference to the base metric on S^2 (which enters in the definition of the mesh of an open cover). One can reformulate (ii) in a similar spirit. If there exists a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$ and $f(\mathcal{C}) \subset \mathcal{C}$, then expansion of the map f can be characterized in yet another way (see Lemma 11.2).

Proof. We will show (i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i) and (iii) \Rightarrow (iv) \Rightarrow (iii).

(i) \Rightarrow (ii): Suppose f is expanding. Pick a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$ and let $\delta_0 > 0$ be as in (7.3) (note that $\#\text{post}(f) \geq 3$ by Corollary 6.4). Suppose \mathcal{U} is a cover of S^2 by open and connected sets that satisfies $\text{mesh}(\mathcal{U}) < \delta_0$. If $U \in \mathcal{U}$, then U is connected and $\text{diam}(U) < \delta_0$. So if V is an arbitrary connected component of $f^{-n}(U)$, then by Lemma 7.8 the set V is contained in an n -flower for (f, \mathcal{C}) . Hence

$$\text{diam}(V) \leq 2 \text{mesh}(f, n, \mathcal{C}),$$

which implies

$$\text{mesh}(f^{-n}(\mathcal{U})) \leq 2 \text{mesh}(f, n, \mathcal{C}).$$

Since f is an expanding Thurston map, we have $\text{mesh}(f, n, \mathcal{C}) \rightarrow 0$, and so $\text{mesh}(f^{-n}(\mathcal{U})) \rightarrow 0$ as $n \rightarrow \infty$.

(ii) \Rightarrow (iii): Obvious.

(iii) \Rightarrow (i): Suppose \mathcal{U} is an open cover of S^2 as in (iii). Pick a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$, and let $\delta > 0$ be a *Lebesgue number* for the cover \mathcal{U} , i.e., every set $K \subset S^2$ with $\text{diam}(K) < \delta$ is contained in a set $U \in \mathcal{U}$. We can find a number $M \in \mathbb{N}$ such that each of the

two 0-tiles for (f, \mathcal{C}) can be written as a union of M connected sets V with $\text{diam}(V) < \delta$. Then each such set V is contained in a set $U \in \mathcal{U}$.

Now if X is an arbitrary n -tile for (f, \mathcal{C}) , then $Y = f^n(X)$ is a 0-tile for (f, \mathcal{C}) and $f^n|X$ is a homeomorphism of X onto Y . Hence X is a union of M connected sets of the form $(f^n|Y)^{-1}(V)$, where $V \subset Y$ is connected and lies in a set $U \in \mathcal{U}$. Then $(f^n|X)^{-1}(V)$ lies in a component of $f^{-n}(U)$, and so

$$\text{diam}((f^n|X)^{-1}(V)) \leq \text{mesh}(f^{-n}(\mathcal{U})).$$

This implies

$$\text{diam}(X) \leq M \text{mesh}(f^{-n}(\mathcal{U})).$$

Hence

$$\text{mesh}(f, n, \mathcal{C}) \leq M \text{mesh}(f^{-n}(\mathcal{U})).$$

Since $\text{mesh}(f^{-n}(\mathcal{U})) \rightarrow 0$, we also have $\text{mesh}(f, n, \mathcal{C}) \rightarrow 0$ as $n \rightarrow \infty$. Hence f is expanding.

(iii) \Rightarrow (iv): Suppose \mathcal{U} is an open cover of S^2 as in (iii), and \mathcal{V} is an arbitrary open cover of S^2 . Let $\delta > 0$ be a Lebesgue number for the cover \mathcal{V} , i.e., every set $K \subset S^2$ with $\text{diam}(K) < \delta$ is contained in a set $V \in \mathcal{V}$. By (iii) we can find $N \in \mathbb{N}$ such that $\text{mesh}(f^{-n}(\mathcal{U})) < \delta$ for $n > N$. If $n > N$ and U' is a set in $f^{-n}(\mathcal{U})$, then $\text{diam}(U') < \delta$ by definition of $\text{mesh}(f^{-n}(\mathcal{U}))$. Hence there exists $V \in \mathcal{V}$ such that $U' \subset V$.

(iv) \Rightarrow (iii): Suppose \mathcal{U} is an open cover of S^2 as in (iv). Then \mathcal{U} also satisfies condition (iii); indeed, let $\epsilon > 0$ be arbitrary, and let \mathcal{V} be the open cover of S^2 consisting of all open balls of radius $\epsilon/2$. Then $\text{diam}(V) \leq \epsilon$ for all $V \in \mathcal{V}$. Moreover, by (iv) there exists $N \in \mathbb{N}$ such that for $n > N$ every set in $f^{-n}(\mathcal{U})$ is contained in a set in \mathcal{V} . In particular, $\text{mesh}(f^{-n}(\mathcal{U})) \leq \epsilon$ for $n > N$. This shows that \mathcal{U} satisfies condition (iii). \square

Let $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$. We define $\tilde{D}_n = \tilde{D}_n(f, \mathcal{C})$ as the minimal number of tiles of order $l \geq n$ for (f, \mathcal{C}) required to join opposite sides of \mathcal{C} , i.e., the smallest number $N \in \mathbb{N}$ for which there are tiles $X_i \in \bigcup_{l \geq n} \mathbf{X}^l$, $i = 1, \dots, N$, such that $K = \bigcup_{i=1}^N X_i$ is connected and joins opposite sides of \mathcal{C} .

Note that the quantity \tilde{D}_n is a variant of the quantity D_n defined in (7.4). While the sets K used to define D_n are unions of tiles of order n , the sets K in the definition of \tilde{D}_n are unions of tiles of order $k \geq n$; in particular, $D_k \geq \tilde{D}_n$ for $k \geq n$.

Lemma 8.3. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$. Let $D_n = D_n(f, \mathcal{C})$ and $\tilde{D}_n = \tilde{D}_n(f, \mathcal{C})$ for $n \in \mathbb{N}_0$.*

Then $D_n \rightarrow \infty$ and $\tilde{D}_n \rightarrow \infty$ as $n \rightarrow \infty$.

Proof. We know that $D_k \geq \tilde{D}_n$ whenever $k \geq n$. So it suffices to show $\tilde{D}_n \rightarrow \infty$ as $n \rightarrow \infty$.

Let $\delta_0 > 0$ be defined as in (7.3) and suppose $K = X_1 \cup \dots \cup X_N$ is a connected union of tiles of order $\geq n$ that joins opposite sides of \mathcal{C} . Then

$$\begin{aligned} \delta_0 &\leq \text{diam}(K) \leq \sum_{i=1}^N \text{diam}(X_i) \\ &\leq N \max_{i=1, \dots, N} \text{diam}(X_i) \\ &\leq N \sup_{k \geq n} \text{mesh}(f, k, \mathcal{C}). \end{aligned}$$

Putting $c_n := \sup_{k \geq n} \text{mesh}(f, k, \mathcal{C})$, we conclude that $N \geq \delta_0/c_n$, and so $\tilde{D}_n \geq \delta_0/c_n$.

Since f is expanding we have $\text{mesh}(f, n, \mathcal{C}) \rightarrow 0$ and so also $c_n \rightarrow 0$ as $n \rightarrow \infty$. This implies that $\tilde{D}_n \rightarrow \infty$ as desired. \square

If f is expanding and \mathcal{C} is given, then in view of the last lemma, we can find a number $k_0 = k_0(f, \mathcal{C}) \in \mathbb{N}$ such that

$$(8.4) \quad \tilde{D}_{k_0} = \tilde{D}_{k_0}(f, \mathcal{C}) \geq 10.$$

This inequality will be useful in the following.

Lemma 8.4. *Let $f: S^2 \rightarrow S^2$ be a Thurston map, $n \in \mathbb{N}$, and $F = f^n$. Then F is a Thurston map with $\text{post}(F) = \text{post}(f)$. The map f is expanding if and only if F is expanding.*

Proof. Since f is a Thurston map, the map F is a branched covering map on S^2 with $\text{post}(F) = \text{post}(f)$ (see Section 3) and $\deg(F) = \deg(f)^n \geq 2$. Hence F is also a Thurston map.

Fix a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) = \text{post}(F) \subset \mathcal{C}$. It follows from the definitions that

$$\text{mesh}(F, k, \mathcal{C}) = \text{mesh}(f, nk, \mathcal{C})$$

for all $k \in \mathbb{N}_0$. If f is expanding, then by Lemma 8.1 we have $\text{mesh}(f, k, \mathcal{C}) \rightarrow 0$ as $k \rightarrow \infty$ which implies that

$$\text{mesh}(F, k, \mathcal{C}) = \text{mesh}(f, nk, \mathcal{C}) \rightarrow 0$$

as $k \rightarrow \infty$. Hence F is expanding.

Conversely, suppose that F is expanding. Then we know that

$$(8.5) \quad \lim_{k \rightarrow \infty} \text{mesh}(F, k, \mathcal{C}) = \lim_{k \rightarrow \infty} \text{mesh}(f, nk, \mathcal{C}) = 0.$$

Let the constant $M \geq 1$ be as in Lemma 7.11 for the map f and the Jordan curve \mathcal{C} . By an argument similar as in the proof of Lemma 8.1 one can show that

$$\text{mesh}(f, l+1, \mathcal{C}) \leq 2M \text{mesh}(f, l, \mathcal{C})$$

for all $l \in \mathbb{N}_0$. This implies

$$\text{mesh}(f, l, \mathcal{C}) \leq (2M)^n \text{mesh}(f, n\lfloor l/n \rfloor, \mathcal{C})$$

for all $l \in \mathbb{N}_0$ and so by (8.5) we have $\text{mesh}(f, l, \mathcal{C}) \rightarrow 0$ as $l \rightarrow \infty$. This shows that f is expanding. \square

Definition 8.5. Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$, and $x, y \in S^2$.

For $x \neq y$ we define

$$m_{f,\mathcal{C}}(x, y) := \max\{n \in \mathbb{N}_0 : \text{there exist non-disjoint } n\text{-tiles } X \text{ and } Y \text{ for } (f, \mathcal{C}) \text{ with } x \in X, y \in Y\}.$$

If $x = y$ we define $m_{f,\mathcal{C}}(x, x) := \infty$.

Note that a maximal number n as in the definition of $m_{f,\mathcal{C}}(x, y)$ for $x \neq y$ exists, because we know that for an expanding Thurston map the diameters of n -tiles tend to 0 if $m \rightarrow \infty$. We usually drop one or both subscripts in $m_{f,\mathcal{C}}(x, y)$ if f or \mathcal{C} are clear from the context. A similar combinatorial quantity that is essentially equivalent to $m_{f,\mathcal{C}}(x, y)$ for $x \neq y$ is

$$m'_{f,\mathcal{C}}(x, y) := \min\{n \in \mathbb{N}_0 : \text{there exist disjoint } n\text{-tiles } X \text{ and } Y \text{ for } (f, \mathcal{C}) \text{ with } x \in X, y \in Y\}$$

(see Lemma 8.6 (v)).

In the next lemma we collect some of the properties of the function $m_{f,\mathcal{C}}$. For the proof the following terminology is useful. A *(finite) chain* in S^2 is a finite sequence A_1, \dots, A_N of sets in S^2 such that $A_i \cap A_{i+1} \neq \emptyset$ for $i = 1, \dots, N-1$. It *joins* two points $x, y \in S^2$ if $x \in A_1$ and $y \in A_N$. We say that this chain is *simple* if there is no proper subsequence of A_1, \dots, A_N that is also chain joining x and y .

If K is a compact connected set in S^2 , \mathcal{U} an open cover of K , and $x, y \in K$, then one can always find a simple chain of sets in \mathcal{U} joining x and y .

Lemma 8.6. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map, $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$, and $m = m_{f,\mathcal{C}}$.*

(i) *There exists a number $k_1 > 0$ such that*

$$(8.6) \quad \min\{m(x, z), m(y, z)\} \leq m(x, y) + k_1$$

for all $x, y, z \in S^2$.

(ii) *We have*

$$m(f(x), f(y)) \geq m(x, y) - 1$$

for all $x, y \in S^2$.

(iii) *Let $\tilde{\mathcal{C}} \subset S^2$ be another Jordan curve with $\text{post}(f) \subset \tilde{\mathcal{C}}$. Then there exists a constant $k_2 > 0$ such that*

$$(8.7) \quad m(x, y) - k_2 \leq m_{f,\tilde{\mathcal{C}}}(x, y) \leq m(x, y) + k_2$$

for all $x, y \in S^2$.

(iv) *Let $F = f^n$ be an iterate of f . Then there exists a constant $k_3 > 0$ such that*

$$m(x, y) - k_3 \leq n \cdot m_{F,\mathcal{C}}(x, y) \leq m(x, y),$$

for all $x, y \in S^2$.

(v) *There exists a constant $k_4 > 0$ with the following property: if $x, y \in S^2$, $x \neq y$, and $m'_{f,\mathcal{C}}(x, y)$ is the smallest number $m' \in \mathbb{N}_0$ for which there exist m' -tiles $X', Y' \in \mathcal{D}^{m'}(f, \mathcal{C})$ with $x \in X'$, $y \in Y'$ and $X' \cap Y' = \emptyset$, then*

$$m(x, y) - k_4 \leq m'_{f,\mathcal{C}}(x, y) \leq m(x, y) + 1.$$

Proof. We fix $k_0 = k_0(f, \mathcal{C})$ as in (8.4). Let $x, y \in S^2$ be arbitrary. In order to establish the desired inequalities we may always assume $x \neq y$. Unless otherwise stated, tiles will be for (f, \mathcal{C}) .

(i) Let $m = m(x, y) \in \mathbb{N}_0$ be as in Definition 8.5. We can pick $(m+1)$ -tiles X_0 and Y_0 containing x and y , respectively. Then $X_0 \cap Y_0 = \emptyset$ by definition of m .

Define $n := m + k_0$, and let $z \in S^2$ be arbitrary. We claim that $m(x, z) \leq n$ or $m(y, z) \leq n$.

Otherwise $m(x, z) \geq n + 1$ and $m(y, z) \geq n + 1$, and so by Definition 8.5 there exist numbers $m_1, m_2 \geq n + 1$ and m_1 -tiles X and Z with $x \in X$, $z \in Z$ and $X \cap Z \neq \emptyset$, and m_2 -tiles Y and Z' with $y \in Y$, $z \in Z'$ and $X \cap Z' \neq \emptyset$.

Then the set $K = X \cup Z \cup Z' \cup Y$ is connected and meets the disjoint $(m+1)$ -tiles X_0 and Y_0 . Thus $f^{m+1}(K)$ joins opposite sides of \mathcal{C} by Lemma 7.9, and consists of four tiles of order $\geq n - m = k_0$. This contradicts (8.4), proving the claim.

So we have $m(x, z) \leq m + k_0$ or $m(y, z) \leq m + k_0$. This implies (8.6) with the constant $k_1 = k_0$ which is independent of x and y .

(ii) We may assume that $m = m(x, y) \geq 1$. Then there are non-disjoint m -tiles X and Y with $x \in X$ and $y \in Y$. It follows that $f(X)$ and $f(Y)$ are non-disjoint $(m - 1)$ -tiles with $f(x) \in f(X)$ and $f(y) \in f(Y)$. Hence $m(f(x), f(y)) \geq m - 1$ as desired.

(iii) Let $\tilde{m} = m_{f, \tilde{\mathcal{C}}}(x, y) \in \mathbb{N}_0$. Then there exist \tilde{m} -tiles \tilde{X} and \tilde{Y} for $(f, \tilde{\mathcal{C}})$ with $x \in \tilde{X}$, $y \in \tilde{Y}$, and $\tilde{X} \cap \tilde{Y} \neq \emptyset$. By Lemma 7.12 the sets \tilde{X} and \tilde{Y} are each contained in M \tilde{m} -flowers for (f, \mathcal{C}) , where M is independent of \tilde{X} and \tilde{Y} . In particular, this implies that we can find a chain of at most M such \tilde{m} -flowers joining x and y . Since any two tiles in the closure $\overline{W^n(v)}$ of an n -flower have the point v in common, it follows that there exists a chain X_1, \dots, X_N of \tilde{m} -tiles for (f, \mathcal{C}) joining x and y with $N \leq 2M$. Let $x_1 := x$, $x_N := y$, and for $i = 2, \dots, N - 1$, pick a point $x_i \in X_i$. Then $m(x_i, x_{i+1}) \geq \tilde{m}$ for $i = 1, \dots, N - 1$. Hence by repeated application of (i) we obtain

$$\begin{aligned} \tilde{m} &\leq \min\{m(x_i, x_{i+1}) : i = 1, \dots, N - 1\} \\ &\leq m(x_1, x_N) + Nk_1 \leq m(x, y) + 2Mk_1. \end{aligned}$$

Since $2Mk_1$ is independent of x and y , we get an upper bound as in (8.7). A lower bound is obtained by the same argument if we reverse the roles of \mathcal{C} and $\tilde{\mathcal{C}}$.

(iv) The map F is also an expanding Thurston map, and we have $\text{post}(f) = \text{post}(F)$ (see Lemma 8.4); so the Jordan curve \mathcal{C} contains the set of postcritical points of F and $m_{F, \mathcal{C}}$ is defined. It follows from Proposition 6.1 (v) that the m -tiles for (f, \mathcal{C}) are precisely the (nm) -tiles for (F, \mathcal{C}) . In the ensuing proof we will only consider tiles for (f, \mathcal{C}) .

Let $m_F = m_{F, \mathcal{C}}(x, y)$ and $m = m(x, y)$; then there are non-disjoint (nm_F) -tiles X and Y with $x \in X$ and $y \in Y$. So $m \geq nm_F$ which gives the desired upper bound.

We claim that on the other hand, we have $m \leq nm_F + k_3$, where $k_3 = n + k_0 - 1$. To see this assume that

$$m \geq nm_F + k_3 + 1 = n(m_F + 1) + k_0.$$

Then we can find non-disjoint m -tiles X and Y with $x \in X$, $y \in Y$. Moreover, we can pick $n(m_F + 1)$ -tiles X' and Y' with $x \in X'$ and $y \in Y'$. By definition of m_F we know that $X' \cap Y' = \emptyset$, so X' and Y' are disjoint $n(m_F + 1)$ -tiles joined by the connected set $K = X \cup Y$. Hence by Lemma 7.10 K must consist of at least

$$D_{m-n(m_F+1)} \geq \tilde{D}_{k_0} \geq 10$$

m -tiles; but K consists of only two such m -tiles. This is a contradiction showing the desired claim.

(v) Let $m' = m'_{f,C}(x, y)$ be defined as in (v). Then $m' \geq 1$, because the two 0-tiles have nonempty intersection. So $m' - 1 \geq 0$, and there exist $(m' - 1)$ -tiles X and Y with $x \in X$ and $y \in Y$. Then $X \cap Y \neq \emptyset$ by definition of m' , and so $m(x, y) \geq m' - 1$.

Conversely, let $m = m(x, y)$. Suppose $m' < m - k_0$. Then there exist m' -tiles X' and Y' with $X' \cap Y' = \emptyset$, m -tiles X and Y with $X \cap Y \neq \emptyset$, and $x \in X \cap X'$, $y \in Y \cap Y'$. Hence $K = X \cup Y$ is a union of two m -tiles joining the disjoint m' -tiles X' and Y' ; but such a union must consist of at least at

$$D_{m-m'} \geq \tilde{D}_{k_0} \geq 10$$

m -tiles by Lemma 7.10. This is a contradiction showing that $m - k_0 \leq m'$. So the claim is true with $k_4 = k_0$. \square

Definition 8.7 (Visual metrics). Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map and d be a metric on S^2 . Then d is called a *visual metric* (for f) if there exists a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$, and constants $\Lambda > 1$, and $A \geq 1$ such that

$$(8.8) \quad (1/A)\Lambda^{-m(x,y)} \leq d(x, y) \leq A\Lambda^{-m(x,y)}$$

for all $x, y \in S^2$, where $m(x, y) = m_{f,C}(x, y)$.

Here we use the convention $\Lambda^{-\infty} = 0$. The number Λ is called the *expansion factor* of the metric d . It is easy to see that the expansion factor of each visual metric is uniquely determined; different visual metrics may have different expansion factors.

Let d and d' be two metrics on a space X . They are called *bi-Lipschitz equivalent* if there exists a constant $C \geq 1$ such that

$$(1/C)d(x, y) \leq d'(x, y) \leq Cd(x, y)$$

for all $x, y \in X$. They are called *snowflake equivalent* if there exist constants $\alpha > 0$ and $C \geq 1$ such that

$$(1/C)d(x, y) \leq d'(x, y)^\alpha \leq Cd(x, y)$$

for all $x, y \in X$. Obviously, bi-Lipschitz equivalence of two metrics implies their snowflake equivalence.

Remark 8.8. As briefly mentioned in the introduction, one can use an expanding Thurston map $f: S^2 \rightarrow S^2$ to define an infinite graph \mathcal{G} that is Gromov hyperbolic (see [GH, BuS] for the definition of Gromov hyperbolic spaces and an explanation of the related terminology employed in the present remark). Namely, we consider a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$ and the tiles in the cell decompositions

$\mathcal{D}^n(f, \mathcal{C})$, $n \in \mathbb{N}_0$. The vertex set of \mathcal{G} is the set of all tiles, and one connects two vertices by an edge in \mathcal{G} if the corresponding tiles have nonempty intersection and levels that differ by at most 1 (there are other reasonable ways to define the edges in \mathcal{G}). One can show that \mathcal{G} is Gromov hyperbolic, and that the boundary at infinity $\partial_\infty \mathcal{G}$ of \mathcal{G} can be identified with S^2 . Moreover, $m_{f, \mathcal{C}}(x, y)$ is essentially the Gromov product of two points $x, y \in S^2 \cong \partial_\infty \mathcal{G}$ (with one of the 0-tiles chosen as a basepoint in \mathcal{G}). Then the notion of a visual metric on $\partial_\infty \mathcal{G}$ as in the theory of Gromov hyperbolic spaces coincides with our notion of a visual metric on $S^2 \cong \partial_\infty \mathcal{G}$.

In the following proposition we summarize properties of visual metrics.

Proposition 8.9. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map.*

- (i) *There exist visual metrics for f .*
- (ii) *Every visual metric induces the standard topology on S^2 .*
- (iii) *Let d be a visual metric with expansion factor Λ . Then an inequality as in (8.8) with the same expansion factor Λ holds for every Jordan curve $\tilde{\mathcal{C}} \subset S^2$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$ if $m = m_{f, \tilde{\mathcal{C}}}$ and A is suitably chosen depending on $\tilde{\mathcal{C}}$.*
- (iv) *Any two visual metrics are snowflake equivalent; if they have the same expansion factor Λ , then they are bi-Lipschitz equivalent.*
- (v) *A metric is a visual metric for any iterate $F = f^n$ if and only if it is a visual metric for f .*

Proof. (i) Fix a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$. Recall that a function $q: S^2 \times S^2 \rightarrow [0, \infty)$ is called a *quasimetric* if it has the symmetry property $q(x, y) = q(y, x)$, satisfies the conditions $q(x, y) = 0 \Leftrightarrow x = y$, and the inequality

$$(8.9) \quad q(x, y) \leq K(q(x, z) + q(z, y)),$$

holds for a constant $K \geq 1$ and all $x, y, z \in S^2$.

We now define a quasimetric q on S^2 . To this purpose, we fix $\Lambda > 1$ and set

$$(8.10) \quad q(x, y) := \Lambda^{-m(x, y)},$$

for $x, y \in S^2$, where $m(x, y) = m_{f, \mathcal{C}}(x, y) \in \mathbb{N}_0 \cup \{\infty\}$ is as in Definition 8.5.

Symmetry and the property $q(x, y) = 0 \Leftrightarrow x = y$ are clear. The quasi-triangle inequality (8.9) follows from Lemma 8.6 (i).

It is well known (see [He, Prop. 14.5]) that a sufficient “snowflaking” of a quasimetric leads to a distance function that is comparable to a metric. This means there is a metric d and $0 < \epsilon < 1$ such that $q^\epsilon \asymp d$. Then d is a visual metric for f (with expansion factor Λ^ϵ).

(ii) Let d be a visual metric for f satisfying (8.8), and d' our fixed “base metric” on S^2 that induces the standard topology of S^2 . We have to show that if $x \in S^2$ and $\{x_i\}$ is a sequence in S^2 , then $d(x_i, x) \rightarrow 0$ if and only if $d'(x_i, x) \rightarrow 0$ as $i \rightarrow \infty$.

Now by (8.8) the relation $d(x_i, x) \rightarrow 0$ is obviously equivalent to $m_i := m_{f, \mathcal{C}}(x_i, x) \rightarrow \infty$.

So if $d(x_i, x) \rightarrow 0$, then $m_i \rightarrow \infty$, and so for each $n \in \mathbb{N}_0$ we have $m_i \geq n$ for sufficiently large i . For these i we have

$$d'(x_i, x) \leq 2 \sup_{k \geq n} \text{mesh}(f, k, \mathcal{C})$$

as follows from the definition of $m_{f, \mathcal{C}}$. Since f is expanding, we know that $\sup_{k \geq n} \text{mesh}(f, k, \mathcal{C}) \rightarrow 0$ as $n \rightarrow \infty$. Hence $d'(x_i, x) \rightarrow 0$ as $i \rightarrow \infty$.

Conversely, suppose that $d'(x_i, x) \rightarrow 0$ as $i \rightarrow \infty$. Let $n \in \mathbb{N}_0$ be arbitrary. Then x lies in some n -flower $W^n(v)$. Since flowers are open sets, we have $x_i \in W^n(v)$ for sufficiently large i . For each of these i we can find n -tiles X and Y with $x \in X$, $x_i \in Y$, and $v \in X \cap Y$. This implies $m_i \geq n$. Hence $m_i \rightarrow \infty$ as desired.

(iii) This follows from Lemma 8.6 (iii).

(iv) This follows from (iii) and the definition of a visual metric.

(v) This follows from (iii) and Lemma 8.6 (iv). \square

The next lemma gives a geometric characterization of visual metrics.

Lemma 8.10. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map, $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$, and d be a visual metric for f with expansion factor $\Lambda > 1$. Then there exists a constant $C \geq 1$ such that*

- (i) $\text{dist}_d(\sigma, \tau) \geq (1/C)\Lambda^{-n}$ whenever σ and τ are disjoint n -cells,
- (ii) $(1/C)\Lambda^{-n} \leq \text{diam}_d(\tau) \leq C\Lambda^{-n}$ for all n -edges and all n -tiles τ .

Conversely, if d is a metric on S^2 satisfying conditions (i) and (ii) for some constant $C \geq 1$, then d is a visual metric with expansion factor $\Lambda > 1$.

Here it is understood that n -cells are for (f, \mathcal{C}) .

Proof. By Proposition 8.9 (iii) we may assume that d satisfies (8.8), where $m = m_{f, \mathcal{C}}$.

(i) Let k_0 be defined as in (8.4), and let σ and τ be disjoint n -cells. If $x \in \sigma$ and $y \in \tau$ are arbitrary, then $m = m(x, y) < n + k_0$. Indeed, if

this were not the case, then we could find $(n+k)$ -tiles X and Y with $x \in X$, $y \in Y$, $X \cap Y \neq \emptyset$, and $k \geq k_0$. Then $K = X \cup Y$ is a connected set meeting disjoint n -cells. Hence by Lemma 7.9 the set $f^n(K)$ joins opposite sides of \mathcal{C} . On the other hand, $f^n(K)$ consists of two k -tiles $X' = f^n(X)$ and $Y' = f^n(Y)$, where $k \geq k_0$. This is impossible by definition of k_0 .

Therefore, $d(x, y) \geq (1/A)\Lambda^{-n-k_0}$, and so we get the desired bound $\text{dist}(\sigma, \tau) \geq (1/C')\Lambda^{-n}$ with the constant $C' = C\Lambda^{k_0}$ that is independent of n , σ , and τ .

(ii) If x, y are points in some n -tile X , then $m(x, y) \geq n$. Since every n -edge is contained in an n -tile, this inequality is still true if x and y are contained in an n -edge. Hence $d(x, y) \leq A\Lambda^{-m(x, y)} \leq A\Lambda^{-n}$, and so $\text{diam}(\tau) \leq A\Lambda^{-n}$ whenever τ is an n -tile or n -edge, where the constant A is as in (8.8).

A similar lower bound for the diameter of an n -edge or n -tile τ follows from (i) and the fact that every n -edge or n -tile contains two distinct n -vertices.

For the converse suppose that we have (i) and (ii). Let $x, y \in S^2$, $x \neq y$, be arbitrary, and $m = m_{f, \mathcal{C}}(x, y)$.

Then we can find m -tiles X and Y with $x \in X$, $y \in Y$ and $X \cap Y \neq \emptyset$. By (ii) we have

$$d(x, y) \leq \text{diam}(X) + \text{diam}(Y) \lesssim \Lambda^{-m}.$$

We can also find $(m+1)$ -tiles X' and Y' with $x \in X'$, $y \in Y'$. By definition of m we then have $X' \cap Y' = \emptyset$. Hence by (i)

$$d(x, y) \geq \text{dist}(X', Y') \gtrsim \Lambda^{-m}.$$

Since the implicit multiplicative constants in the previous inequalities are independent of x and y , it follows that d is a visual metric. \square

It is possible to establish this phenomenon of “exponential shrinking” for other types of sets. For example, we have

$$\text{diam}(W^n(v)) \leq C\Lambda^{-n}$$

for every n -flower for (f, \mathcal{C}) where the constant C is independent of n and v . Of particular importance will be exponential shrinking for lifts of paths.

Lemma 8.11. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map, and d be a visual metric for f with expansion factor $\Lambda > 1$. Then for every path $\gamma: [0, 1] \rightarrow S^2$ there exists a constant $A \geq 1$ with the following property: if $\tilde{\gamma}$ is any lift of γ under f^n , then*

$$\text{diam}_d(\tilde{\gamma}) \leq A\Lambda^{-n}.$$

Proof. Pick a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$, and let $\delta_0 > 0$ be as in (7.3). Then we can break up γ into a finite number of paths γ_i , $i = 1, \dots, N$, traversed in successive order such that $\text{diam}(\gamma_i) < \delta_0$ for all $i = 1, \dots, N$. By Lemma 7.8 (ii) each lift of the pieces γ_i is contained in one n -flower, and so the whole lift $\tilde{\gamma}$ in N n -flowers. Hence by Lemma 8.10 we have $\text{diam}(\tilde{\gamma}) \leq C N \Lambda^{-n}$ with a constant C independent of n and γ . \square

In general the constant A in the last lemma will dependent on γ , but the proof shows that we can take the same constant A for a family of paths if there exists $N \in \mathbb{N}$ such that each path can be broken up into at most N subpaths of diameter $< \delta_0$.

Let f be an expanding Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$. It is useful to define neighborhoods of points by using the cells in our decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$. To define them let $x \in S^2$ and $n \in \mathbb{N}_0$, and set

$$(8.11) \quad U^n(x) = \bigcup \{Y \in \mathbf{X}^n : \text{there exists an } n\text{-tile } X \text{ with} \\ x \in X \text{ and } X \cap Y \neq \emptyset\}.$$

It is convenient to define $U^n(x)$ also for negative integers n . We set $U^n(x) = U^0(x) = S^2$ for $n < 0$. It follows from Lemma 8.10 that the sets $U^n(x)$ resemble metric balls very closely.

Lemma 8.12. *Let d be a visual metric for f with expansion factor $\Lambda > 1$. Then there are constants $K \geq 1$ and $n_0 \in \mathbb{N}_0$ with the following properties.*

(i) *For all $x \in S^2$ and all $n \in \mathbb{Z}$*

$$B_d(x, r/K) \subset U^n(x) \subset B_d(x, Kr),$$

where $r = \Lambda^{-n}$.

(ii) *For all $x \in S^2$ and all $r > 0$*

$$U^{n+n_0}(x) \subset B_d(x, r) \subset U^{n-n_0}(x),$$

where $n = \lceil -\log r / \log \Lambda \rceil$.

Proof. (i) Let $m = m_{f, \mathcal{C}}$. If $y \in U^n(x)$, then $m(x, y) \geq n$, and so $d(x, y) \lesssim \Lambda^{-n} = r$. This gives the inclusion $U^n(x) \subset B_d(x, Kr)$ for a suitable constant K independent of x and n .

Conversely, suppose that $y \notin U^n(x)$. Then $n \geq 1$. If we pick n -tiles X and Y with $x \in X$ and $y \in Y$, then $X \cap Y = \emptyset$ by definition of $U^n(x)$. So by Lemma 8.10 (ii) we have

$$d(x, y) \geq \text{dist}(X, Y) \gtrsim \Lambda^{-n} = r.$$

Hence $B_d(x, r/K) \subset U^n(x)$ if K is suitably large independent of x and r .

(ii) Choose $n_0 = \lceil \log K / \log \Lambda \rceil + 1$, where K is as in (i). Then $\Lambda^{-n_0} \leq 1/(\Lambda K)$. Moreover, $\Lambda^{-n} \leq r \leq \Lambda \Lambda^{-n}$, and so

$$K \Lambda^{-n-n_0} \leq r \leq (1/K) \Lambda^{-n+n_0}.$$

The desired inclusion then follows from (i). \square

Lemma 8.13. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map, $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$, and d be a visual metric for f with expansion factor $\Lambda > 1$. Then there exists a constant $C \geq 1$ with the following property: for every n -tile X for (f, \mathcal{C}) there exists a point $p \in X$ such that*

$$B_d(p, (1/C) \Lambda^{-n}) \subset X \subset B_d(p, C \Lambda^{-n}).$$

So if S^2 is equipped with a visual metric, then tiles are “quasi-round”, and every tile contains points that are “deep inside” the tile.

Proof. With a suitable constant C independent of n , an inclusion of the form

$$X \subset B_d(p, C \Lambda^{-n})$$

holds for every n -tile X and every point $p \in X$ as follows from Lemma 8.10 (i).

The main difficulty for an inclusion in the opposite direction is to find an appropriate point p . For this purpose let $k_0 \in \mathbb{N}$ be the number defined in (8.4), and X be an arbitrary n -tile. Since f is an expanding Thurston map, we have $\text{post}(f) \geq 3$ (see Corollary 6.4), and so ∂X contains at least three distinct n -vertices v_1, v_2, v_3 . Using these vertices, we can find three arcs $\alpha_1, \alpha_2, \alpha_3 \subset \partial X$ with pairwise disjoint interior such that $\partial X = \alpha_1 \cup \alpha_2 \cup \alpha_3$ and such that α_i has the endpoints v_i and v_{i+1} for $i = 1, 2, 3$, where $v_4 = v_1$. In general α_i will not be an n -edge, but since it lies on ∂X , and its endpoints are n -vertices, it is the union of all the n -edges that it contains.

We now define

$$A_i = \bigcup_{x \in \alpha_i} U^{n+k_0}(x)$$

for $i = 1, 2, 3$, where $U^{n+k_0}(x)$ is defined as in (8.11). Then the set A_i is the union of all $(n+k_0)$ -tiles that meet an $(n+k_0)$ -tile that has nonempty intersection with α_i . In particular, A_i is a closed set that contains α_i .

We claim that the sets A_1, A_2, A_3 do not form a cover of X . To reach a contradiction suppose that $X \subset A_1 \cup A_2 \cup A_3$. We can regard X as a topological simplex with the sides α_i , $i = 1, 2, 3$. Then the closed sets

A_1, A_2, A_3 form a cover of X such that each set A_i contains the side α_i of the simplex for $i = 1, 2, 3$. A well-known result due to Sperner [AH, p. 378] then implies that $A_1 \cap A_2 \cap A_3 \neq \emptyset$.

Pick a point $x \in A_1 \cap A_2 \cap A_3$. Then by definition of A_i , there exist $(n + k_0)$ -tiles X_i and Y_i with $X_i \cap \alpha_i \neq \emptyset$, $x \in Y_i$, and $X_i \cap Y_i \neq \emptyset$, where $i = 1, 2, 3$. Then the set

$$K = \bigcup_{i=1}^3 (X_i \cup Y_i)$$

consists of at most six $(n + k_0)$ -tiles, is connected, and meets each of the arcs $\alpha_1, \alpha_2, \alpha_3$. Hence $K' = f^n(K)$ is a connected set that consists of at most six k_0 -tiles, and meets each of the arcs $\beta_i = f^n(\alpha_i)$, $i = 1, 2, 3$. Note that each arc β_i is the union of all 0-edges edges that it contains. Hence for $i = 1, 2, 3$ there exists a 0-edge $e_i \subset \beta_i$ with $e_i \cap K' \neq \emptyset$. Since the arcs $\beta_1, \beta_2, \beta_3$ have pairwise disjoint interior, it follows that the 0-edges e_1, e_2, e_3 are all distinct. So K' is a connected set that meets three distinct 0-edges. Hence it joins opposite sides of \mathcal{C} . So K' should contain at least $D_{k_0} \geq \tilde{D}_{k_0} \geq 10$ tiles of order k_0 . This is a contradiction, because K' is a union of at most six k_0 -tiles.

This proves the claim that the sets A_1, A_2, A_3 do not cover X , and we conclude that we can find a point

$$p \in X \setminus (A_1 \cup A_2 \cup A_3).$$

We claim that $U^{n+k_0}(p) \subset X$. If not, we could find a point $y \in U^{n+k_0}(p) \setminus X$, and $(n + k_0)$ -tiles U and V with $p \in U$, $y \in V$, and $U \cap V \neq \emptyset$. Then the connected set $U \cup V$ must meet ∂X , and hence one of the arcs α_i ; but then $p \in A_i$ by definition of A_i . This is a contradiction showing the desired inclusion $U^{n+k_0}(p) \subset X$.

Using Lemma 8.12 (i) it follows that $B_d(p, (1/C)\Lambda^{-n}) \subset X$, where $C \geq 1$ is a constant independent of n and X . \square

9. SYMBOLIC DYNAMICS

Shift operators serve as important paradigms in symbolic dynamics. Often a goal in understanding a dynamical system (X, f) given by the iteration of a map f on a space X is to link it to shift operators, or more generally, to shifts of finite type. For expanding Thurston maps this is accomplished by Theorem 1.6 stated in the introduction. The statement is essentially due to Kameyama (see [Ka, Thm. 3.4]). His notion of an expanding Thurston map is different from ours, but his proof carries over to our setting with only minor modifications (see

below). In this section we will also establish a related fact (see Proposition 9.1) for maps with cellular Markov partitions as introduced in Section 4. We start with some basic definitions.

Let J be a finite set. We consider J as an *alphabet* and its elements as *letters* in this alphabet. A *word* is a finite sequence $w = i_1 \dots i_n$, where $n \in \mathbb{N}_0$ and $i_1, \dots, i_n \in J$. For $n = 0$ we interpret this as the *empty word* \emptyset . The number n is called the *length* of the word $w = i_1 i_2 \dots i_n$. The words of length n can be identified with n -tuples in J , i.e., elements of the Cartesian power J^n . The letters, i.e., the elements in J , are precisely the words of length 1. If $w = i_1 i_2 \dots i_n$ and $w' = j_1 \dots j_m$, then we denote by $ww' = i_1 i_2 \dots i_n j_1 \dots j_m$ the word obtained by concatenating w and w' .

Let J^* be the set of all words in the alphabet J . The *(left-)shift* $\Sigma: J^* \rightarrow J^*$ is defined by setting $\Sigma(i_1 i_2 \dots i_n) = i_2 \dots i_n$ for a word $w = i_1 i_2 \dots i_n \in J^*$. We denote by J^ω the set of all sequences (i_n) in J , where the sequence elements $i_n \in J$ are indexed by $n \in \mathbb{N}$. More informally, we consider a sequence $s = (i_n) \in J^\omega$ as a “word of infinite length” and write $s = i_1 i_2 \dots$. If $s = (i_n) \in J^\omega$ and $n \in \mathbb{N}_0$, then we denote by $[s]_n \in J^*$ the word $s_n = i_1 \dots i_n$ consisting of the first n elements of the sequence s . The *(left-)shift* $\Sigma: J^\omega \rightarrow J^\omega$ is the map that assigns to each sequence $(i_n) \in J^\omega$ the sequence $(j_n) \in J^\omega$ with $j_n = i_{n+1}$ for all $n \in \mathbb{N}$. In our notation we do not distinguish the shifts on J^* and J^ω and denote both maps by Σ . Note that $[\Sigma(s)]_n = \Sigma([s]_{n+1})$ for all $s \in J^\omega$; indeed, if $s = i_1 i_2 \dots$, then we have

$$[\Sigma(s)]_n = [i_2 i_3 \dots]_n = i_2 \dots i_{n+1} = \Sigma(i_1 \dots i_{n+1}) = \Sigma([s]_{n+1}).$$

If we equip J with the discrete topology, then J^ω carries a natural metrizable product topology. This topology is induced by the ultrametric d given by $d(s, s') = 2^{-N}$ for $s = (i_n) \in J^\omega$ and $s' = (j_n) \in J^\omega$, $s \neq s'$, where $N = \min\{n : i_n \neq j_n\}$. In particular, two elements $s, s' \in J^\omega$ are close if and only if $s_n = s'_n$ for some large n . Equipped with this topology, the space J^ω is compact.

Let $T: J \times J \rightarrow \{0, 1\}$ be a map encoding “allowed” transitions between the letters, and let J_T^ω be the set of all sequences (i_n) in J^ω such that $T(i_n, i_{n+1}) = 1$ for all $n \in \mathbb{N}$. Then $\Sigma(J_T^\omega) \subset J_T^\omega$, and so we can consider Σ as a map on J_T^ω . A *subshift of finite type* is a map of the form $\Sigma_T := \Sigma|_{J_T^\omega}$ for some finite set J and some map $T: J \times J \rightarrow \{0, 1\}$.

Suppose that X and \tilde{X} are topological spaces, and $f: X \rightarrow X$ and $\tilde{f}: \tilde{X} \rightarrow \tilde{X}$ are continuous maps. We say that the dynamical system (X, f) is a *factor* of the dynamical system

(\tilde{X}, \tilde{f}) if there exists a surjective continuous map $\varphi: \tilde{X} \rightarrow X$ such that $\varphi \circ \tilde{f} = f \circ \varphi$.

The following proposition shows that under some mild additional assumptions a map with a cellular Markov partition can be obtained as a factor of a subshift of finite type. Since we know that at least some iterate f^n of an expanding Thurston map f has a cellular Markov partition as in this proposition, it follows immediately that f^n is a factor of a subshift of finite type. Of course, by Theorem 1.6, which will be proved below, an even stronger statement is true in this case.

Proposition 9.1. *Let (X, d) be a compact metric space, $f: X \rightarrow X$ be a continuous map with a cellular Markov partition $(\mathcal{D}', \mathcal{D})$, and \mathcal{D}^n for $n \in \mathbb{N}_0$ be the cell decompositions of X as given by Proposition 4.10. Suppose that*

$$\lim_{n \rightarrow \infty} \max_{\tau \in \mathcal{D}^n} \text{diam}(\tau) = 0.$$

Then f is a factor of a subshift of finite type.

Proof. Let $J = \mathcal{D}^1 = \mathcal{D}'$. Note that since X is compact, the set J is finite. As above we let J^ω be the set of all sequences in J , and $\Sigma: J^\omega \rightarrow J^\omega$ be the shift operator. Define the map $T: J \times J \rightarrow \{0, 1\}$ as follows: if $\sigma, \tau \in J = \mathcal{D}^1$, we put

$$T(\sigma, \tau) = 1 \text{ if } \tau \subset f(\sigma) \text{ and } T(\sigma, \tau) = 0 \text{ otherwise.}$$

So we have $T(\sigma, \tau) = 1$ precisely if $\tau \in \mathcal{D}^1$ is one of the cells into which $f(\sigma) \in \mathcal{D}^0$ is subdivided.

We want to prove that (X, f) is a factor of (J_T^ω, Σ_T) , where $\Sigma_T = \Sigma|J_T^\omega$.

Let \mathcal{S} be the set of all sequences (σ_n) , where $\sigma_n \in \mathcal{D}^n$ and $\sigma_{n+1} \subset \sigma_n$ for $n \in \mathbb{N}$. Since f^{n-1} is cellular for $(\mathcal{D}^n, \mathcal{D}^1)$, it follows that if $(\sigma_n) \in \mathcal{S}$, then $(x_n) \in J^\omega$, where $x_n = f^{n-1}(\sigma_n) \in \mathcal{D}^1 = J$ for $n \in \mathbb{N}$. Moreover, for each $n \in \mathbb{N}$ we have $x_{n+1} = f^n(\sigma_{n+1}) \subset f^n(\sigma_n) = f(x_n)$, and so $T(x_n, x_{n+1}) = 1$. It follows that $(x_n) \in J_T^\omega$. In this way we get a map

$$\Phi: \mathcal{S} \rightarrow J_T^\omega, \quad (\sigma_n) \in \mathcal{S} \mapsto \Phi[(\sigma_n)] := (f^{n-1}(\sigma_n)).$$

The map Φ is a bijection. To show injectivity, suppose that (σ_n) and (τ_n) are sequences in \mathcal{S} , and $\Phi[(\sigma_n)] = \Phi[(\tau_n)]$. Then $f^{n-1}(\sigma_n) = f^{n-1}(\tau_n)$ for all $n \in \mathbb{N}$. We show inductively that this implies $\sigma_n = \tau_n$ for all $n \in \mathbb{N}$. Indeed, for $n = 1$ we have $\sigma_1 = \tau_1$ by definition of Φ . Suppose that $\sigma_n = \tau_n =: \lambda$. Since $f^n|_\lambda$ is a homeomorphism of λ onto $f^n(\lambda) \in \mathcal{D}^0$, and $\sigma_{n+1}, \tau_{n+1} \subset \lambda$, we have

$$\sigma_{n+1} = (f^n|_\lambda)^{-1}(f^n(\sigma_{n+1})) = (f^n|_\lambda)^{-1}(f^n(\tau_{n+1})) = \tau_{n+1}.$$

Hence $(\sigma_n) = (\tau_n)$.

To show surjectivity of Φ , let $(x_n) \in J_T^\omega$ be arbitrary. We define a sequence of cells $\sigma_n \in \mathcal{D}^n$ with $f^{n-1}(\sigma_n) = x_n$ inductively as follows. Let $\sigma_1 := x_1 \in \mathcal{D}^1$. Suppose σ_n is already defined such that $x_n = f^{n-1}(\sigma_n)$. Since $T(x_n, x_{n+1}) = 1$, we have $x_{n+1} \subset f(x_n)$. Pick a point $q \in \text{int}(x_{n+1})$. Since $f^n|_{\sigma_n}$ is a homeomorphism of σ_n onto $f(x_n) \in \mathcal{D}^0$, there exists a unique point $p \in \sigma_n$ with $f^n(p) = q$. Since \mathcal{D}^{n+1} is a refinement of \mathcal{D}^n , the cells in \mathcal{D}^{n+1} contained in σ_n form a cell decomposition of σ_n . Hence σ_n is the disjoint union of the interiors of the cells in \mathcal{D}^{n+1} contained in σ_n . So there exists a unique cell $\tau \in \mathcal{D}^{n+1}$ with $\tau \subset \sigma_n$ and $p \in \text{int}(\tau)$. Then $f^n(\tau)$ is a cell in \mathcal{D}^1 with $q = f^n(p) \in \text{int}(f^n(\tau))$. Since x_{n+1} is the unique cell in \mathcal{D}^1 containing q in its interior, we must have $x_{n+1} = f^n(\tau)$. Now define $\sigma_{n+1} := \tau$. Then $f^n(\sigma_{n+1}) = x_{n+1}$ as desired. Note that by construction $\sigma_{n+1} \subset \sigma_n$ for $n \in \mathbb{N}$. Hence $(\sigma_n) \in \mathcal{S}$, and we have $\Phi[(\sigma_n)] = (x_n)$.

We define a map $\Psi: \mathcal{S} \rightarrow X$ as follows: if $(\sigma_n) \in \mathcal{S}$, then $\sigma_{n+1} \subset \sigma_n$ for $n \in \mathbb{N}$ and $\text{diam}(\sigma_n) \rightarrow 0$ as $n \rightarrow \infty$ by our hypotheses. Hence the intersection $\bigcap_{n \in \mathbb{N}} \sigma_n$ contains a unique point $p \in X$. Set $\Psi[(\sigma_n)] = p$.

Now let $\varphi := \Psi \circ \Phi^{-1}: J_T^\omega \rightarrow X$. Then this map is continuous on J_T^ω . We sketch the proof for this, leaving the details, which can easily be filled in, to the reader. If (x_n) and (y_n) are points in J_T^ω that are “close”, then there exists large $N \in \mathbb{N}$ such that $x_1 = y_1, \dots, x_N = y_N$. If $(\sigma_n) = \Phi^{-1}[(x_n)]$ and $(\tau_n) = \Phi^{-1}[(y_n)]$, then the argument used for establishing the injectivity of Φ shows that $\sigma_1 = \tau_1, \dots, \sigma_N = \tau_N$. Hence if $p = \varphi[(x_n)]$ and $q = \varphi[(y_n)]$, then $p, q \in \sigma_N = \tau_N \in \mathcal{D}^N$. By our hypotheses the diameter of a cell in \mathcal{D}^N is small if N is large. Hence p and q are close if (x_n) and (y_n) are close. The continuity of φ follows.

The map φ is surjective. To see this let $p \in X$ be arbitrary. By the bijectivity of Φ it is enough to find a sequence $(\sigma_n) \in \mathcal{S}$ with $p \in \bigcap_{n \in \mathbb{N}} \sigma_n$. Appropriate cells σ_n can be found as follows: Since the cells in \mathcal{D}^1 cover X , there exists $\sigma_1 \in \mathcal{D}^1$ with $p \in \sigma_1$. Since \mathcal{D}^2 is a refinement of \mathcal{D}^1 , the cells in \mathcal{D}^2 contained in σ_1 cover σ_1 . Hence there exists a cell $\sigma_2 \in \mathcal{D}^2$ with $p \in \sigma_2$ and $\sigma_2 \subset \sigma_1$. Repeating this argument for σ_2 , we can find a cell $\sigma_3 \in \mathcal{D}^3$ with $\sigma_3 \subset \sigma_2$, and $p \in \sigma_3$, etc. In this way we get a sequence (σ_n) as desired.

Finally, in order to show that $\varphi \circ \Sigma_T = f \circ \varphi = f \circ \Psi \circ \Phi^{-1}$, let $(x_n) \in \Sigma_T$ be arbitrary, $(\sigma_n) = \Phi^{-1}[(x_n)]$, and $p = \varphi[(x_n)]$ be the unique point in the intersection $\bigcap_{n \in \mathbb{N}} \sigma_n$. Define $\tau_n = f(\sigma_{n+1})$ for $n \in \mathbb{N}$. Then $\tau_n \in \mathcal{D}^n$ and $\tau_{n+1} \subset \tau_n$ for $n \in \mathbb{N}$. Hence $(\tau_n) \in \mathcal{S}$. If $(y_n) = \Phi[(\tau_n)] \in J_T^\omega$, then $y_n = f^{n-1}(\tau_n) = f^n(\sigma_{n+1}) = x_{n+1}$. Hence $(y_n) = S_T[(x_n)]$. Moreover, $f(p) \in \bigcap_{n \in \mathbb{N}} \tau_n$, and so $f(p) = \varphi[(y_n)] =$

$(\varphi \circ \Sigma_T)[(x_n)]$. On the other hand, $f(p) = (f \circ \varphi)[(x_n)]$. The desired identity $\varphi \circ \Sigma_T = f \circ \varphi$ follows. We conclude that (X, f) is a factor of (J_T^ω, Σ_T) , and hence a factor of a subshift of finite type. \square

Remark 9.2. Let $\varphi: J_T^\omega \rightarrow X$ be as in the proof of the last proposition. If we define an equivalence relation \sim on J_T^ω by

$$(x_n) \sim (y_n) \iff \varphi[(x_n)] = \varphi[(y_n)]$$

for $(x_n), (y_n) \in J_T^\omega$, then the quotient space $\tilde{X} := J_T^\omega / \sim$ is homeomorphic to X and the map $\tilde{f}: \tilde{X} \rightarrow \tilde{X}$ induced by Σ_T on \tilde{X} is topologically conjugate to $f: X \rightarrow X$.

We are now ready to prove Theorem 1.6.

Proof of Theorem 1.6. Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map, and $k := \deg(f) \geq 2$. Fix a visual metric d for f , and let $\Lambda > 1$ be its expansion factor. In the following metric concepts refer to d . We consider tiles for (f, \mathcal{C}) and color them black and white as in Lemma 6.2. Choose a basepoint $p \in S^2 \setminus \text{post}(f)$ in the interior of the white 0-tile X_w^0 .

Claim 1. $\sup_{x \in S^2} \text{dist}(x, f^{-n}(p)) \lesssim \Lambda^{-n}$,

where $C(\lesssim)$ is independent of n . In other words, the set $f^{-n}(p)$ forms a very dense net in S^2 if n is large.

To see this let $x \in S^2$ be arbitrary. Then x lies in some n -tile X^n . If X^n is white, then X^n contains a point in $f^{-n}(p)$ and so $\text{dist}(x, f^{-n}(p)) \leq \text{diam}(X^n)$. If X^n is black, then X^n shares an edge with a white n -tile Y^n . Then Y^n contains a point in $f^{-n}(p)$, and so $\text{dist}(x, f^{-n}(p)) \leq \text{diam}(X^n) + \text{diam}(Y^n)$.

From the inequalities in both cases and Lemma 8.10 we conclude

$$\text{dist}(x, f^{-n}(p)) \lesssim \Lambda^{-n},$$

where $C(\lesssim)$ is independent of x and n . Claim 1 follows.

None of the points in $S^2 \setminus \text{post}(f)$ is a critical value for any of the iterates f^n of f . Moreover, each iterate f^n is a covering map $f^n: S^2 \setminus f^{-n}(\text{post}(f)) \rightarrow S^2 \setminus \text{post}(f)$. Since $p \in S^2 \setminus \text{post}(f)$, we have $f^{-n}(p) \subset S^2 \setminus f^{-n}(\text{post}(f))$ and

$$(9.1) \quad \#f^{-n}(p) = \deg(f^n) = \deg(f)^n = k^n$$

for $n \in \mathbb{N}$. In particular,

$$f^{-1}(p) \subset S^2 \setminus f^{-1}(\text{post}(f)) \subset S^2 \setminus \text{post}(f),$$

and $\#f^{-1}(p) = k$. Let $q_1, \dots, q_k \in S^2 \setminus \text{post}(f)$ be the points in $f^{-1}(p)$. For $i = 1, \dots, k$ we pick a path $\alpha_i: [0, 1] \rightarrow S^2 \setminus \text{post}(f)$ with $\alpha_i(0) = p$ and $\alpha_i(1) = q_i$.

Let $J := \{1, \dots, k\}$, and consider the shift $\Sigma: J^\omega \rightarrow J^\omega$. We want to show that f is a factor of Σ , i.e., that there exists a continuous and surjective map $\varphi: J^\omega \rightarrow S^2$ with $f \circ \varphi = \varphi \circ \Sigma$. In order to define φ , we first construct a suitable map ψ that assigns to each word in J^* a point in S^2 .

Definition of ψ . The map $\psi: J^* \rightarrow S^2$ will be defined inductively such that

$$\psi(w) \in f^{-n}(p),$$

whenever $n \in \mathbb{N}_0$ and $w \in J^n \subset J^*$ is a word of length n . For the empty word \emptyset we set $\psi(\emptyset) = p$, and for the word consisting of the single letter $i \in J$ we set $\psi(i) := q_i \in f^{-1}(p)$.

Now suppose that ψ has been defined for all words of length $\leq n$, where $n \in \mathbb{N}$. Let w be an arbitrary word of length $n+1$. Then $w = w'i$, where $w' \in J^*$ is a word of length n and $i \in J$. So $\psi(w') \in f^{-n}(p)$ is already defined. Since $f^n(\psi(w')) = p$ and $f^n: S^2 \setminus f^{-n}(\text{post}(f)) \rightarrow S^2 \setminus \text{post}(f)$ is a covering map, the path α_i has a unique lift with initial point $\psi(w')$, i.e., there exists a unique path $\tilde{\alpha}_i: [0, 1] \rightarrow S^2$ with $\tilde{\alpha}_i(0) = \psi(w')$ and $f^n \circ \tilde{\alpha}_i = \alpha_i$. We now define $\psi(w) := \tilde{\alpha}_i(1)$. Note that then

$$f^{n+1}(\psi(w)) = f^{n+1}(\tilde{\alpha}_i(1)) = f(\alpha_i(1)) = f(q_i) = p.$$

Hence $\psi(w) \in f^{-(n+1)}(p)$. This shows that a map $\psi: J^* \rightarrow S^2$ with the desired properties exists.

Claim 2. $f(\psi(w)) = \psi(\Sigma(w))$ for all non-empty words $w \in J^*$.

We prove this by induction on the length of the word w . If $w = i \in J$, then

$$f(\psi(w)) = f(\psi(i)) = f(q_i) = p = \psi(\emptyset) = \psi(\Sigma(i)) = \psi(\Sigma(w)).$$

So the claim is true for words of length 1.

Suppose the claim is true for words of length $\leq n$, where $n \in \mathbb{N}$. Let w be a word of length $n+1$. Then $w = w'i$, where w' is a word of length n and $i \in J$. Let $\tilde{\alpha}_i$ be the path as above used in the definition of $\psi(w)$. Define $\tilde{\beta}_i := f \circ \tilde{\alpha}_i$. Then $\tilde{\beta}_i$ is a lift of α_i by f^{n-1} . By induction hypothesis its initial point is

$$\tilde{\beta}_i(0) = f(\tilde{\alpha}_i(0)) = f(\psi(w')) = \psi(\Sigma(w')).$$

In other words, $\tilde{\beta}_i$ is the unique path as in the definition of ψ used to determine $\psi(\Sigma(w')i)$ from $\psi(\Sigma(w'))$, i.e., $\psi(\Sigma(w')i) = \tilde{\beta}_i(1)$. Hence

$$\psi(\Sigma(w)) = \psi(\Sigma(w')i) = \tilde{\beta}_i(1) = f(\tilde{\alpha}_i(1)) = f(\psi(w'i)) = f(\psi(w))$$

as desired, and Claim 2 follows.

Claim 3. For each $n \in \mathbb{N}$ the map $\psi|J^n: J^n \rightarrow f^{-n}(p)$ is a bijection. In other words, the map ψ provides a “coding” of the points in $f^{-n}(p)$ by words of length n .

Again we prove this by induction on n . By definition of ψ it is true for $n = 1$.

Suppose it is true for some $n \in \mathbb{N}$. It suffices to show that the map $\psi|J^{n+1}: J^{n+1} \rightarrow f^{-(n+1)}(p)$ is surjective, since both sets J^{n+1} and $f^{-(n+1)}(p)$ have the same cardinality k^{n+1} . So let $x \in f^{-(n+1)}(p)$ be arbitrary. Then $f^n(x) \in f^{-1}(p)$, and so there exists $i \in J$ with $f^n(p) = q_i$. Since

$$x \in f^{-(n+1)}(p) \subset S^2 \setminus f^{-(n+1)}(\text{post}(f)) \subset S^2 \setminus f^{-n}(\text{post}(f)),$$

and $f^n: S^2 \setminus f^{-n}(\text{post}(f)) \rightarrow S^2 \setminus \text{post}(f)$ is a covering map, we can lift the path α_i by f^n to a path $\tilde{\alpha}_i: [0, 1] \rightarrow S^2$ whose terminal point is x (to see this, lift α_i traversed in opposite direction so that the initial point of the lift is x). Then $f^n(\tilde{\alpha}_i(0)) = \alpha_i(0) = p$, and so $\tilde{\alpha}_i(0) \in f^{-n}(p)$. By induction hypothesis there exists a word $w' \in J^n$ with $\psi(w') = \tilde{\alpha}_i(0)$. Then $\tilde{\alpha}_i$ is a path as used to determine $\psi(w'i)$ from $\psi(w')$. So if we set $w := w'i \in J^{n+1}$, then

$$\psi(w) = \psi(w'i) = \tilde{\alpha}_i(1) = x.$$

This shows that $\psi|J^{n+1}: J^{n+1} \rightarrow f^{-(n+1)}(p)$ is surjective. Claim 3 follows.

Claim 4. If $s \in J^\omega$, then the points $\psi([s]_n)$, $n \in \mathbb{N}$, form a Cauchy sequence in S^2 (recall that $[s]_n$ is the word consisting of the first n elements of the sequence s).

By definition of ψ the points $\psi([s]_n)$ and $\psi([s]_{n+1})$ are joined by a lift of one of the paths $\alpha_1, \dots, \alpha_k$ by f^n . Hence by Lemma 8.11 we have

$$(9.2) \quad d(\psi([s]_n), \psi([s]_{n+1})) \lesssim \Lambda^{-n},$$

where $C(\lesssim)$ is independent of n and s . Hence $(\psi([s]_n))$ is a Cauchy sequence.

Definition of φ . If $s \in J^\omega$, then by Claim 4 the limit

$$\varphi(s) := \lim_{n \rightarrow \infty} \psi([s]_n)$$

exists. This defines a map $\varphi: J^\omega \rightarrow S^2$.

Claim 5. $f \circ \varphi = \varphi \circ \Sigma$.

To see this, let $s \in J^\omega$ be arbitrary. Note that $\Sigma([s]_n) = [\Sigma(s)]_{n-1}$ for $n \in \mathbb{N}$. Hence by Claim 2 and the continuity of f we have

$$f(\varphi(s)) = \lim_{n \rightarrow \infty} f(\psi([s]_n)) = \lim_{n \rightarrow \infty} \psi(\Sigma([s]_n)) = \lim_{n \rightarrow \infty} \psi([\Sigma(s)]_{n-1}) = \varphi(\Sigma(s)).$$

The claim follows.

Claim 6. The map $\varphi: J^\omega \rightarrow S^2$ is continuous and surjective.

Let $s \in J^\omega$ and $n \in \mathbb{N}$. Then (9.2) shows that

$$(9.3) \quad d(\varphi(s), \psi([s]_n)) \lesssim \sum_{l=n}^{\infty} \Lambda^{-l} \lesssim \Lambda^{-n},$$

where $C(\lesssim)$ is independent of n and s . Hence if $s, s' \in J^\omega$ and $[s]_n = [s']_n$, then

$$d(\varphi(s), \varphi(s')) \lesssim \Lambda^{-n},$$

where $C(\lesssim)$ is independent of n , s , and s' . The continuity of φ follows from this; indeed, if s and s' are “close” in J^ω , then $[s]_n = [s']_n$ for some large n , and so the image points $\varphi(s)$ and $\varphi(s')$ are close in S^2 .

Since J^ω is compact, the continuity of φ implies that the image $\varphi(J^\omega)$ is also compact and hence closed in S^2 . The surjectivity of φ will follow, if we can show that φ has dense image in S^2 .

To see this let $x \in S^2$ and $n \in \mathbb{N}$ be arbitrary. Then by Claim 1 we can find a point $y \in f^{-n}(p)$ with $d(x, y) \lesssim \Lambda^{-n}$, where $C(\lesssim)$ is independent of x and n . Moreover, by Claim 3 there exists a word $w \in J^n$ with $\psi(w) = y$. Pick $s \in J^\omega$ such that $[s]_n = w$. Then by (9.3) we have

$$d(x, \varphi(s)) \leq d(x, y) + d(y, \varphi(s)) = d(x, y) + d(\psi([s]_n), \varphi(s)) \lesssim \Lambda^{-n},$$

where $C(\lesssim)$ is independent of the choices. Hence

$$\sup_{x \in S^2} \text{dist}(x, \varphi(J^\omega)) \lesssim \Lambda^{-n}$$

for all n , where $C(\lesssim)$ is independent of n . This shows that $\varphi(J^\omega)$ is dense in S^2 and the claim follows.

The theorem now follows from Claim 5 and Claim 6. \square

The procedure that we employed to code the elements in $f^{-n}(p)$ by words of length n is well-known [Ne, Sect. 5.2].

It is a standard fact in Complex Dynamics that the repelling periodic points of a rational map on $\widehat{\mathbb{C}}$ are dense in its Julia set. The following statement is an analog of this for expanding Thurston maps. As we will see, it easily follows from the proof of Theorem 1.6.

Corollary 9.3. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map. Then the periodic points of f are dense in S^2 .*

Proof. We use the notation and setup of the proof of Theorem 1.6.

It suffices to show that if $x \in S^2$ and $n \in \mathbb{N}$ are arbitrary, then there exists a point $z \in S^2$ with $f^n(z) = z$ and $d(x, z) \lesssim \Lambda^{-n}$. Here and in the following, $C(\lesssim)$ is independent of x and n .

To find such a point z , we apply Claim 1 in the proof of Theorem 1.6 and conclude that there exists $y \in f^{-n}(p)$ with $d(x, y) \lesssim \Lambda^{-n}$. By Claim 3 in this proof there exists a word $w \in J^*$ of length n such that $\psi(w) = y$. Let s be the unique sequence obtained by periodic repetition of the letters in w , i.e., $s \in J^\omega$ is the unique sequence with $[s]_n = w$ and $\Sigma^n(s) = s$. Put $z := \varphi(s)$. Then Claim 5 in the proof of Theorem 1.6 implies

$$f^n(z) = f^n(\varphi(s)) = \varphi(\Sigma^n(s)) = \varphi(s) = z.$$

Moreover, by (9.3) we have

$$d(y, z) = d(\psi(w), \varphi(s)) = d(\psi([s]_n), \varphi(s)) \lesssim \Lambda^{-n},$$

and so

$$d(x, z) \leq d(x, y) + d(y, z) \lesssim \Lambda^{-n}.$$

The statement follows. \square

10. ISOTOPIES

In this section we present some topological facts about isotopies that will be important throughout the paper.

Let $I = [0, 1]$, and X and Y be topological spaces. Recall (see Section 3) that an isotopy between X and Y is a continuous map $H: X \times I \rightarrow Y$ such that each map $H_t := H(\cdot, t)$ is a homeomorphism of X onto Y . If $A \subset X$, then H is an isotopy *relative to* A or *rel. A* if $H_t(a) = H_0(a)$ for all $a \in A$ and $t \in I$. If $\varphi: X \rightarrow Y$ and $\psi: X \rightarrow Y$ are homeomorphisms, we say that φ and ψ are *isotopic (rel. $A \subset X$)* if there exists an isotopy $H: X \times I \rightarrow Y$ (rel. A) such that $H_0 = \varphi$ and $H_1 = \psi$.

When we say that a family H_t (where it is understood that $t \in I$) of homeomorphisms from X onto Y is an isotopy between X and Y , we consider t as a variable in I and mean that the map $(x, t) \in X \times I \mapsto H_t(x)$ is an isotopy. This is a slightly imprecise, but convenient way of expression. If $X = Y$ then H_t is called an isotopy on X . If $A, B, C \subset X$, then we say that B is *isotopic to C rel. A* or *B can be isotoped into C rel. A* if there exists an isotopy $H: X \times I \rightarrow X$ rel. A

with $H_0 = \text{id}_X$ and $H_1(B) = C$. Note that this notion depends on the ambient space X containing the sets A, B, C .

10.1. Equivalent expanding Thurston maps are topologically conjugate. Recall that two Thurston maps $f: S^2 \rightarrow S^2$ and $g: \widehat{S}^2 \rightarrow \widehat{S}^2$ on 2-spheres S^2 and \widehat{S}^2 are (Thurston) equivalent (see Definition 3.3) if there exist homeomorphisms $h_0, h_1: S^2 \rightarrow \widehat{S}^2$ that are isotopic rel. $\text{post}(f)$ and satisfy that $h_0 \circ f = g \circ h_1$. We then have the commutative diagram:

$$(10.1) \quad \begin{array}{ccc} S^2 & \xrightarrow{h_1} & \widehat{S}^2 \\ f \downarrow & & \downarrow g \\ S^2 & \xrightarrow{h_0} & \widehat{S}^2. \end{array}$$

The maps f and g are topologically conjugate if there exists a homeomorphism $h: S^2 \rightarrow \widehat{S}^2$ such that $h \circ f = g \circ h$.

Obviously, the notion of Thurston equivalence is weaker than topological conjugacy of the maps. We will show that under the additional assumption that the maps are expanding, we can promote an equivalence between two Thurston maps to a topological conjugacy. The idea for the proof of this statement uses well-known ideas in dynamics. A statement very similar to Theorem 10.4 below was proved by Kameyama [Ka]. Since his notion of “expanding” is different from ours, we will present the details of the proof.

First, we state a lifting theorem that will be needed (see [Ka, Lem. 4.3]).

Proposition 10.1 (Isotopy lifting for branched covers). *Let $f: S^2 \rightarrow S^2$ and $g: \widehat{S}^2 \rightarrow \widehat{S}^2$ be Thurston maps, and $h_0, \tilde{h}_0: S^2 \rightarrow \widehat{S}^2$ be homeomorphisms such that $h_0|_{\text{post}(f)} = \tilde{h}_0|_{\text{post}(f)}$ and $h_0 \circ f = g \circ \tilde{h}_0$. Suppose $H: S^2 \times I \rightarrow \widehat{S}^2$ is an isotopy rel. $\text{post}(f)$ with $H_0 = h_0$. Then the isotopy H uniquely lifts to an isotopy $\tilde{H}: S^2 \times I \rightarrow \widehat{S}^2$ rel. $f^{-1}(\text{post}(f))$ such that $\tilde{H}_0 = \tilde{h}_0$ and $g \circ \tilde{H}_t = H_t \circ f$ for all $t \in I$.*

So if we set $h_1 := H_1$ and $\tilde{h}_1 := \tilde{H}_1$, then we have the following commutative diagram:

$$\begin{array}{ccc} S^2 & \xrightarrow{\tilde{H}: \tilde{h}_0 \simeq \tilde{h}_1} & \widehat{S}^2 \\ f \downarrow & & \downarrow g \\ S^2 & \xrightarrow{H: h_0 \simeq h_1} & \widehat{S}^2. \end{array}$$

Proof. By an argument similar to the one used to establish (3.4) one can show that

$$(10.2) \quad h_0(\text{post}(f)) = \tilde{h}_0(\text{post}(f)) = \text{post}(g).$$

This implies that

$$H_t(\text{post}(f)) = \text{post}(g)$$

for all $t \in I$. Hence $H_t|_{S^2 \setminus \text{post}(f)}$ is an isotopy between $S^2 \setminus \text{post}(f)$ and $\hat{S}^2 \setminus \text{post}(g)$.

Moreover, it follows from (10.2) that

$$\tilde{h}_0(f^{-1}(\text{post}(f))) = g^{-1}(\text{post}(g)).$$

So the map $\tilde{h}_0|_{S^2 \setminus f^{-1}(\text{post}(f))}$ can be considered as a lift of

$$H_0|_{S^2 \setminus \text{post}(f)} = h_0|_{S^2 \setminus \text{post}(f)}$$

by the (non-branched) covering maps

$$f: S^2 \setminus f^{-1}(\text{post}(f)) \rightarrow S^2 \setminus \text{post}(f)$$

and

$$g: \hat{S}^2 \setminus g^{-1}(\text{post}(g)) \rightarrow \hat{S}^2 \setminus \text{post}(g).$$

By the usual homotopy lifting theorem for covering maps (see [Ha, p. 60, Prop. 1.30]) the isotopy $H_t|_{S^2 \setminus \text{post}(f)}$ lifts to a unique isotopy \tilde{H}_t between $S^2 \setminus f^{-1}(\text{post}(f))$ and $\hat{S}^2 \setminus g^{-1}(\text{post}(g))$ such that

$$\tilde{H}_0 = \tilde{h}_0|_{S^2 \setminus f^{-1}(\text{post}(f))}$$

and $g \circ \tilde{H}_t = H_t \circ f$ on $S^2 \setminus f^{-1}(\text{post}(f))$ for all $t \in I$.

Since H_t is constant in t on $\text{post}(f)$, each map \tilde{H}_t has a continuous extension to S^2 , also denoted by \tilde{H}_t . Then $\tilde{H}_t|_{f^{-1}(\text{post}(f))}$ does not depend on t . Moreover, each map \tilde{H}_t is a homeomorphism from S^2 onto \hat{S}^2 , because an inverse of \tilde{H}_t can be obtained by lifting the isotopy H_t^{-1} . So \tilde{H}_t is an isotopy between S^2 and \hat{S}^2 rel. $f^{-1}(\text{post}(f))$. It is clear that it has the desired properties. \square

Note that if in the previous proposition H is an isotopy relative to a set $M \subset S^2$ with $\text{post}(f) \subset M$, then the lift \tilde{H} is an isotopy rel. $f^{-1}(M)$. Indeed, if $p \in f^{-1}(M)$, then $f(p) \in M$ and so

$$g(\tilde{H}_t(p)) = H_t(f(p)) = h_0(f(p)) =: q$$

for all $t \in I$. Thus $t \mapsto \tilde{H}_t(p)$ is a path contained in the finite set $g^{-1}(q)$ and hence a constant path.

For later use we record a simple lemma about preimages of sets.

Lemma 10.2. *Let $f: X \rightarrow X$ and $g: Y \rightarrow Y$ be maps defined on some sets X and Y , and $h_0, h_1: X \rightarrow Y$ be bijections with $g \circ h_1 = h_0 \circ f$. Then for every set $A \subset X$ we have*

$$g^{-1}(h_0(A)) = h_1(f^{-1}(A)).$$

Proof. Under the given assumptions, consider a set $A \subset X$, and let $y \in g^{-1}(h_0(A)) \subset Y$ be arbitrary. Since h_1 is a bijection, there exists $x \in X$ with $h_1(x) = y$. Then

$$h_0(f(x)) = g(h_1(x)) = g(y) \in h_0(A),$$

and so, since $h_0: X \rightarrow Y$ is also a bijection, we have $f(x) \in A$. This implies that $x \in f^{-1}(A)$ and $y = h_1(x) \in h_1(f^{-1}(A))$. Thus $g^{-1}(h_0(A)) \subset h_1(f^{-1}(A))$.

For the other inclusion, let $y \in h_1(f^{-1}(A))$ be arbitrary. Then

$$g(y) \in (g \circ h_1)(f^{-1}(A)) = (h_0 \circ f)(f^{-1}(A)) \subset h_0(A),$$

and so $y \in g^{-1}(h_0(A))$. Thus, $h_1(f^{-1}(A)) \subset g^{-1}(h_0(A))$, and the claim follows. \square

The following lemma will be of crucial importance.

Lemma 10.3 (Exponential shrinking of tracks of isotopies).

Let $f: S^2 \rightarrow S^2$ and $g: \widehat{S}^2 \rightarrow \widehat{S}^2$ be Thurston maps, and $H^n: S^2 \times I \rightarrow \widehat{S}^2$ be isotopies rel. $\text{post}(f)$ satisfying $g \circ H_t^{n+1} = H_t^n \circ f$ for $n \in \mathbb{N}_0$ and $t \in I$.

If g is expanding and \widehat{S}^2 is equipped with a visual metric for g , then the tracks of the isotopies H^n shrink exponentially as $n \rightarrow \infty$; more precisely, if d is a visual metric for g with expansion factor $\Lambda > 1$, then there exists a constant $C \geq 1$ such that

$$(10.3) \quad \sup_{x \in S^2} \text{diam}_d(\{H_t^n(x) : t \in I\}) \leq C\Lambda^{-n}$$

for all $n \in \mathbb{N}_0$.

Proof. For all $n \in \mathbb{N}_0$ and $t \in I$ we have $g^n \circ H_t^n = H_t^0 \circ f^n$; so for fixed $x \in S^2$ and $n \in \mathbb{N}_0$ the path $t \mapsto H_t^n(x)$ in \widehat{S}^2 is a lift of the path $t \mapsto H_t^0(f^n(x))$ by the map g^n . Recall that in the proof of Lemma 8.11 we had to break up the path γ into N pieces γ_j so that $\text{diam}(\gamma_j) < \delta_0$ (see also (7.3)). Since H^0 is uniformly continuous we can choose the number N uniformly for all the paths $t \mapsto H_t^0(y)$, $y \in S^2$. Since g is expanding, Lemma 8.11 then implies that

$$\sup_{x \in S^2} \text{diam}_d(\{H_t^n(x) : t \in I\}) \lesssim \Lambda^{-n}$$

for all $n \in \mathbb{N}$, where $C(\lesssim)$ is independent of n . \square

Theorem 10.4 (Thurston equivalence implies topological conjugacy). *Let $f: S^2 \rightarrow S^2$ and $g: \widehat{S}^2 \rightarrow \widehat{S}^2$ be equivalent Thurston maps that are expanding. Then they are topologically conjugate. More precisely, if we have a Thurston equivalence between f and g as in (10.1), then there exists a homeomorphism $h: S^2 \rightarrow \widehat{S}^2$ such that h is isotopic to h_1 rel. $f^{-1}(\text{post}(f))$ and satisfies $h \circ f = g \circ h$.*

Since $\text{post}(f) \subset f^{-1}(\text{post}(f))$ and h_0 and h_1 are isotopic rel. $\text{post}(f)$ this implies that h is also isotopic to h_0 rel. $\text{post}(f)$.

Proof. The main idea of the proof is to lift a suitable initial isotopy repeatedly and use the fact that by Lemma 10.3 the tracks of the isotopies shrink exponentially fast. The desired conjugacy is then obtained as a limit.

By assumption there exists an isotopy H_t^0 between S^2 and \widehat{S}^2 rel. $\text{post}(f)$ such that $h_0 \circ f = g \circ h_1$, where $h_0 = H_0^0$ and $h_1 = H_1^0$. By Proposition 10.1 we can lift the isotopy H_t^0 between h_0 and h_1 to an isotopy H_t^1 rel. $f^{-1}(\text{post}(f)) \supset \text{post}(f)$ between h_1 and $h_2 := H_1^1$. Note that the map h_1 plays two roles here: it is the endpoint H_1^0 of the initial isotopy H_t^0 , and also a lift of h_0 .

Repeating this argument we get homeomorphisms h_n and isotopies H_t^n between S^2 and \widehat{S}^2 rel. $f^{-n}(\text{post}(f)) \supset \text{post}(f)$ for all $n \in \mathbb{N}_0$ such that $H_t^n \circ f = g \circ H_t^{n+1}$, $H_0^n = h_n$ and $H_1^n = h_{n+1}$ for all $n \in \mathbb{N}_0$ and $t \in I$. So we have the following “infinite tower” of isotopies:

$$\begin{array}{ccc}
 & \vdots & \\
 \downarrow & & \downarrow \\
 S^2 & \xrightarrow{H^2: h_2 \simeq h_3} & \widehat{S}^2 \\
 \downarrow f & & \downarrow g \\
 S^2 & \xrightarrow{H^1: h_1 \simeq h_2} & \widehat{S}^2 \\
 \downarrow f & & \downarrow g \\
 S^2 & \xrightarrow{H^0: h_0 \simeq h_1} & \widehat{S}^2
 \end{array}$$

We want to show that for $n \rightarrow \infty$ the maps h_n converge to a homeomorphism h_∞ that gives the desired topological conjugacy between f and g .

To see this fix a visual metric d on \widehat{S}^2 , and assume that it has the expansion factor $\Lambda > 1$. Metric concepts on \widehat{S}^2 will refer to this metric in the following. Since g is expanding, Lemma 10.3 implies that

$$(10.4) \quad \sup_{x \in S^2} \text{diam}(\{H_t^n(x) : t \in I\}) \lesssim \Lambda^{-n}$$

for all $n \in \mathbb{N}$, where $C(\lesssim)$ is independent of n . In particular,

$$\text{dist}(h_{n+1}, h_n) := \sup_{x \in S^2} \text{dist}(h_n(x), h_{n+1}(x)) \lesssim \Lambda^{-n}$$

for all $n \in \mathbb{N}_0$, and so there is a continuous map $h_\infty : S^2 \rightarrow \widehat{S}^2$ such that $h_n \rightarrow h_\infty$ uniformly on S^2 as $n \rightarrow \infty$. Since $h_{n-1} \circ f = g \circ h_n$, we have $h_\infty \circ f = g \circ h_\infty$.

The map h_∞ is a homeomorphism. To see this we repeat the argument where we interchange the roles of f and g . More precisely, we consider the isotopy $(H_t^0)^{-1}$ between h_0^{-1} and h_1^{-1} . The corresponding tower of repeated lifts of this initial isotopy is given by the isotopies $(H_t^n)^{-1}$ between h_n^{-1} and h_{n+1}^{-1} . By the argument in the first part of the proof we see that the maps h_n^{-1} converge to a continuous map $k_\infty : \widehat{S}^2 \rightarrow S^2$ uniformly on \widehat{S}^2 as $n \rightarrow \infty$. By uniform convergence we have $k_\infty \circ h_\infty(x) = \lim_{n \rightarrow \infty} h_n^{-1} \circ h_n(x) = x$ for all $x \in S^2$. Hence $k_\infty \circ h_\infty = \text{id}_{S^2}$. Similarly, $h_\infty \circ k_\infty = \text{id}_{\widehat{S}^2}$, and so k_∞ is a continuous inverse of h_∞ . Hence h_∞ is a homeomorphism.

The conjugating map $h = h_\infty$ is isotopic to h_1 rel. $f^{-1}(\text{post}(f))$. To see this we will define an isotopy rel. $f^{-1}(\text{post}(f))$ that is obtained by concatenating (with suitable time change) the isotopies H^1, H^2, \dots and take $h = h_\infty$ as the endpoint at time $t = 1$. The precise definition is as follows. We break up the unit interval into intervals

$$I = [0, 1] = \left[0, \frac{1}{2}\right] \cup \left[\frac{1}{2}, \frac{3}{4}\right] \cup \dots \cup [1 - 2^{-n}, 1 - 2^{-n-1}] \cup \dots \cup \{1\}.$$

The n -th interval in this union is denoted by $I^n = [1 - 2^{-n}, 1 - 2^{-n-1}]$. Let $s_n : I^n \rightarrow I$, $s_n(t) = 2^{n+1}(t - (1 - 2^{-n}))$, for $n \in \mathbb{N}_0$. We define $H : S^2 \times I \rightarrow \widehat{S}^2$ by

$$H(x, t) := H^{n+1}(p, s_n(t))$$

if $p \in S^2$ and $t \in I^n$ for some $n \in \mathbb{N}_0$, and $H(p, t) = h(p)$ for $p \in S^2$ and $t = 1$. We claim that H is indeed an isotopy between h_1 and h rel. $f^{-1}(\text{post}(f))$.

Note that H is well defined, $H_1 = h$, and $H_{1-1/2^n} = h_{n+1}$ for $n \in \mathbb{N}_0$. Moreover, H_t is a homeomorphism for each $t \in I$, and $H_t|_{f^{-1}(\text{post}(f))}$ does not depend on t . To establish our claim, it remains to verify that H is continuous. It is clear that H is continuous at each point $(p, t) \in S^2 \times [0, 1)$.

Moreover, as follows from the uniform convergence $h_n \rightarrow h$ as $n \rightarrow \infty$ and inequality (10.4), we have $H_t \rightarrow H_1$ uniformly on S^2 as $t \rightarrow 1$. This together with the continuity of $h = H_1$ implies the continuity of H at points $(p, t) \in S^2 \times I$ with $t = 1$. \square

Remark 10.5. The previous proof gives a procedure for approximating the conjugating map $h = h_\infty$. Indeed, as follows from the remark after the proof of Proposition 10.1, the map H_t^n is constant in t on $f^{-n}(\text{post}(f))$ for all $n \in \mathbb{N}_0$. This implies that $h_n = h_{n+1} = \dots = h_\infty$ on the set $f^{-n}(\text{post}(f))$, and so the map h_n sends the points in $f^{-n}(\text{post}(f))$ to the “right” points in $g^{-n}(\text{post}(g))$. The isotopy H_t^n then deforms h_n to a map h_{n+1} such that the points in $f^{-(n+1)}(\text{post}(f))$ have the correct images in $g^{-(n+1)}(\text{post}(g))$ as well, etc. Since by expansion the union of the sets

$$\text{post}(f) \subset f^{-1}(\text{post}(f)) \subset f^{-2}(\text{post}(f)) \subset \dots$$

is dense in S^2 , this gives better and better approximations of limit map h_∞ .

To record an immediate consequence of Theorem 10.4, we introduce some terminology related to the notion of snowflake equivalent metrics defined in Section 8. Let (X, d_X) and (Y, d_Y) be metric spaces. A homeomorphism $h: X \rightarrow Y$ is called a *snowflake equivalence* if there exist constants $\alpha > 0$ and $C \geq 1$ such that

$$\frac{1}{C}d_X(x, x')^\alpha \leq d_Y(h(x), h(x')) \leq Cd_X(x, x')^\alpha$$

for all $x, x' \in X$. The spaces X and Y are called *snowflake equivalent* if there exists a snowflake equivalence between X and Y . Note that two metrics d and d' on a space X are snowflake equivalent as defined in Section 8 if and only if the identity map $\text{id}_X: (X, d) \rightarrow (X, d')$ is a snowflake equivalence.

Corollary 10.6. *Let $f: S^2 \rightarrow S^2$ and $g: \widehat{S}^2 \rightarrow \widehat{S}^2$ be expanding Thurston maps that are Thurston equivalent. Then S^2 equipped with any visual metric with respect to f is snowflake equivalent to \widehat{S}^2 equipped with any visual metric with respect to g . Every homeomorphism $h: S^2 \rightarrow \widehat{S}^2$ satisfying $h \circ f = g \circ h$ is a snowflake equivalence.*

Proof. By Theorem 10.4 we know that there exists a topological conjugacy between f and g , i.e., a homeomorphism $h: S^2 \rightarrow \widehat{S}^2$ such that $h \circ f = g \circ h$. Let d be a visual metric on S^2 with respect to f , and \widehat{d} be a visual metric on \widehat{S}^2 with respect to g . Let $\Lambda > 1$ and $\widehat{\Lambda} > 1$ be the expansion factors of d and \widehat{d} , respectively. It suffices to show that $h: (S^2, d) \rightarrow (\widehat{S}^2, \widehat{d})$ is a snowflake equivalence.

To see this pick a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$. Then $\widehat{\mathcal{C}} = h(\mathcal{C})$ is a Jordan curve in \widehat{S}^2 with $\text{post}(g) = h(\text{post}(f)) \subset \widehat{\mathcal{C}}$. Since h conjugates f and g , it follows from Proposition 6.1 (iii) and (v) or, alternatively, from the uniqueness statement in Lemma 5.4 that for each $n \in \mathbb{N}_0$ the images of the cells in the cell decomposition $\mathcal{D}^n := \mathcal{D}^n(f, \mathcal{C})$ of S^2 under the map h are precisely the cells in the cell decomposition $\widehat{\mathcal{D}}^n = \mathcal{D}^n(g, \widehat{\mathcal{C}})$ of \widehat{S}^2 ; so we have

$$(10.5) \quad \widehat{\mathcal{D}}^n = \{h(c) : c \in \mathcal{D}^n\}$$

for all $n \in \mathbb{N}_0$. This implies that

$$\widehat{m}(h(x), h(x')) = m(x, x')$$

for all $x, x' \in S^2$, where $\widehat{m} = m_{g, \widehat{\mathcal{C}}}$ and $m = m_{f, \mathcal{C}}$ (recall Definition 8.5). Combining this with Proposition 8.9 (iii) we see that

$$\widehat{d}(h(x), h(x')) \asymp \widehat{\Lambda}^{-\widehat{m}(h(x), h(x'))} = \widehat{\Lambda}^{-m(x, x')} = \Lambda^{-\alpha m(x, x')} \asymp d(x, x')^\alpha$$

for all $x, x' \in S^2$, where $\alpha = \log(\widehat{\Lambda})/\log(\Lambda)$ and the implicit multiplicative constants do not depend on x and x' . It follows that h is a snowflake equivalence. \square

10.2. Isotopies of Jordan curves. In the following S^2 is a 2-sphere equipped with a fixed base metric. It will be the ambient space for all isotopies. In this subsection we study the problem when two Jordan curves J and K on S^2 passing through a given finite set P of points in the same order can be deformed into each other by an isotopy of S^2 rel. P . If $\#P \leq 3$ this is always the case (see Lemma 10.10 below).

For $\#P \geq 4$ this is not always true as the example in Figure 3 shows. Here $K = S^1$ is the unit circle and $P = \{1, \mathbf{i}, -1, -\mathbf{i}\} \subset S^1$. The Jordan curve J (which contains P) is drawn with a thick line. The curves $K = S^1$ and J are not isotopic rel. P . In fact, J may be obtained from S^1 by a “Dehn twist” around the points $-\mathbf{i}$ and 1 . Note that in this example we can make the Hausdorff distance (see (15.3)) between J and S^1 arbitrarily small.

We will need the following statement.

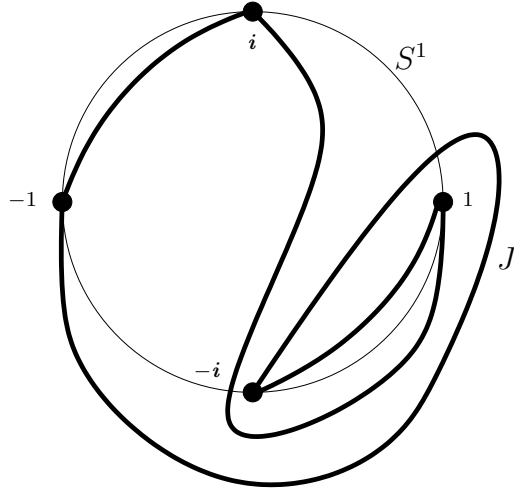


FIGURE 3. J is not isotopic to S^1 rel. $\{1, i, -1, -i\}$.

Proposition 10.7. *Suppose J is a Jordan curve in S^2 and $P \subset J$ a set consisting of $n \geq 3$ distinct points $p_1, \dots, p_n, p_{n+1} = p_1$ in cyclic order on J . For $n = 1, \dots, n$ let α_i be the unique arc on J with endpoints p_i and p_{i+1} such that $\text{int}(\alpha_i) \subset J \setminus P$. Then there exists $\delta > 0$ with the following property:*

Let K be another Jordan curve in S^2 passing through the points p_1, \dots, p_n in cyclic order, and let β_i for $i = 1, \dots, n$ be the arc with endpoints p_i and p_{i+1} such that $\text{int}(\beta_i) \subset J \setminus P$. If

$$\beta_i \subset \mathcal{N}^\delta(\alpha_i)$$

for all $i = 1, \dots, n$, then there exists an isotopy H_t on S^2 rel. P such that $H_0 = \text{id}_{S^2}$ and $H_1(J) = K$.

In other words, if the arcs β_i of the Jordan curve K are contained in sufficiently small neighborhoods of the corresponding arcs α_i of J , then one can deform J into K by an isotopy of S^2 that keeps the points in P fixed. Even though this statement seems “obvious”, a complete proof is surprisingly difficult and involved.

As we will see, the proof of this statement easily follows from two lemmas in [Bu].

Lemma 10.8. *Let $\Omega \subset S^2$ be a simply connected region, $p, q \in \Omega$ distinct points, and α and β arcs in Ω with endpoints p and q . Then α is isotopic to β rel. $\{p, q\} \cup S^2 \setminus \Omega$.*

So arcs in a simply connected region with the same endpoints can be deformed into each other so that the endpoints and the complement

of the region stay fixed. The lemma follows from [Bu, p. 413, A.6 Thm. (ii)].

Lemma 10.9. *Suppose we have two Jordan curves J and K as in Proposition 10.7 such that for each $i = 1, \dots, n$ the arc α_i is isotopic to β_i rel. P .*

Then J is isotopic to K rel. P .

This is essentially [Bu, p. 411, A.5 Thm.].

Proof of Proposition 10.7. For each arc α_i there exists a simply connected region Ω_i that contains α_i but does not contain any element of P different from the endpoints of α_i . There exists $\delta > 0$ such that $\mathcal{N}^\delta(\alpha_i) \subset \Omega_i$ for all $i = 1, \dots, n$. Then by Lemma 10.8 every arc β_i in $\mathcal{N}^\delta(\alpha_i)$ with the same endpoints as α_i can be isotoped to α_i rel. P . The proposition now follows from Lemma 10.9. \square

If $\#P \leq 3$ in Proposition 10.7, then J can always be isotoped to K rel. P .

Lemma 10.10. *Suppose J and K are Jordan curves in S^2 and $P \subset J \cap K$ is a set with $\#P \leq 3$. Then J is isotopic to K rel. P .*

Proof. Suppose first that P consists of exactly three distinct points p_1, p_2, p_3 . Define the arcs α_i and β_i as in Proposition 10.7. Then for each $i = 1, 2, 3$ the arcs α_i and β_i have the same endpoints p_i and p_{i+1} and are contained in the simply connected region $\Omega_i = S^2 \setminus \{p_{i+2}\}$, where indices are understood modulo 3. Hence by Lemma 10.8 each arc α_i is isotopic to β_i rel. P . Again Lemma 10.9 implies that J is isotopic to K rel. P .

If $\#P \leq 2$, we may assume that $S^2 = \widehat{\mathbb{C}}$. Then by applying the first part of the proof (by adding auxiliary points to P) one sees that both J and K are isotopic to circles on $\widehat{\mathbb{C}}$ rel. P . Hence J is isotopic to K rel. P . \square

Lemma 10.11. *Let S^2 and \widehat{S}^2 be oriented 2-spheres, and $P \subset S^2$ be a set with $\#P \leq 3$. If $\alpha: S^2 \rightarrow \widehat{S}^2$ and $\beta: S^2 \rightarrow \widehat{S}^2$ are orientation-preserving homeomorphisms with $\alpha|_P = \beta|_P$, then α and β are isotopic rel. P .*

Proof. The statement is essentially well-known. For the sake of completeness we will give a proof, but will leave some of the details to the reader. These details can easily be filled in by using the facts about isotopies that will be discussed later before Proposition 12.3. In the proof of the uniqueness part of this proposition, we will use very similar arguments.

By considering $\alpha \circ \beta^{-1}$ one can reduce the lemma to the case where $S^2 = \widehat{S}^2$ and $\beta = \text{id}_{S^2}$. Then α fixes the points in P , and we have to show that α is isotopic to id_{S^2} rel. P . We first assume that $\#P = 3$.

Pick a Jordan curve $K \subset S^2$ with $P \subset K$, and let $J = \alpha(K)$. Then $P \subset J \cap K$, and so by Lemma 10.10 the Jordan curve J can be isotoped into K rel. P . This implies that α is isotopic rel. P to a homeomorphism α_1 on S^2 with $\alpha_1(K) = K$. Let e be one of the three subarcs of K determined by P . Since α_1 fixes the three points in P , this map restricts to a homeomorphism of e that does not move the endpoints of e . Hence on e the map α_1 is isotopic to the identity on e rel. ∂e . By pasting the isotopies on these arcs together, one can find an isotopy $h: K \times I \rightarrow K$ rel. P such that $h_0 = \alpha_1|_K$ and $h_1 = \text{id}_K$. One can extend h to each of the complementary components of K to obtain an isotopy $H: S^2 \times I \rightarrow S^2$ rel. P such that $H_1 = \text{id}_{S^2}$ and $H(p, t) = h(p, t)$ for all $p \in K$ and $t \in I$. Then $\alpha_2 := H_0$ is a homeomorphism on S^2 that is isotopic to id_{S^2} rel. P such that $\alpha_1|_K = \alpha_2|_K$. This implies that α_1 and α_2 are isotopic rel. $K \supset P$. If \sim indicates that two homeomorphisms on S^2 are isotopic rel. P , then we have $\alpha \sim \alpha_1 \sim \alpha_2 \sim \text{id}_{S^2}$, and so $\alpha \sim \text{id}_{S^2}$ as desired.

If $\#P \leq 2$, then we pick a set $P' \subset S^2$ with $\#P' = 3$ and $P' \supset P$. By the first part of the proof it suffices to find an isotopy rel. P of the given map α to a homeomorphism α' that fixes the points in P' . It is clear that such an isotopy can always be found; for an explicit construction one can assume that $S^2 = \widehat{\mathbb{C}}$ and can then obtain the desired isotopy by post-composing α by a suitable continuous family of Möbius transformations, for example. \square

The following lemma will be crucial for the proof of the uniqueness statement on invariant Jordan curves. In its proof we will use the following topological fact: if D is a two-dimensional cell and $\varphi: D \rightarrow S^2$ is a continuous map such that $\varphi|_{\partial D}$ is injective, then the set $\varphi(\text{int}(D))$ contains one of the two complementary components of the Jordan curve $\varphi(\partial D)$. Indeed, by applying the Schönflies Theorem and using auxiliary homeomorphisms we can reduce to the case where $D = \overline{\mathbb{D}}$, $S^2 = \widehat{\mathbb{C}}$, $\varphi|_{\partial \mathbb{D}} = \text{id}_{\partial \mathbb{D}}$, and $\infty \notin \varphi(D)$. Then $\mathbb{D} \subset \varphi(\mathbb{D})$. This follows from a simple degree argument and the statement can be generalized to higher dimensions; for an elementary exposition of this and related facts in dimension 2 see [Bur], in particular [Bur, Cor. 3.5].

Lemma 10.12. *Let \mathcal{D} be a cell decomposition of S^2 with 1-skeleton E and vertex set \mathbf{V} , and suppose that every tile in \mathcal{D} contains at least*

three vertices in its boundary. If J and K are Jordan curves that are both contained in E and are isotopic rel. \mathbf{V} , then $J = K$.

Proof. Let $H: S^2 \times I \rightarrow S^2$ be an isotopy rel. \mathbf{V} such that $H_0 = \text{id}_{S^2}$ and $H_1(J) = K$.

Note that if $M \subset S^2$ is a set disjoint from \mathbf{V} , then it remains disjoint from \mathbf{V} during the isotopy, i.e., if $M \cap \mathbf{V} = \emptyset$, then $H_t(M) \cap \mathbf{V} = \emptyset$ for all $t \in I$. This follows from the fact that each map H_t , $t \in [0, 1]$, is a homeomorphism on S^2 with $H_t|_{\mathbf{V}} = \text{id}_{\mathbf{V}}$.

Let e be an edge in \mathcal{D} . We claim that if $H_1(e) \subset E$, then $H_1(e) = e$. First note that $H_1(e)$ is an edge in \mathcal{D} . Indeed, since $\partial e \subset \mathbf{V}$ and the isotopy H does not move vertices, the arc $H_1(e)$ has the same endpoints as e . Moreover, $\text{int}(e) \cap \mathbf{V} = \emptyset$, and so $H_1(\text{int}(e)) \cap \mathbf{V} = \emptyset$ by what we have just seen. So $H_1(\text{int}(e))$ is a connected set in the 1-skeleton E of \mathcal{D} disjoint from the 0-skeleton \mathbf{V} . By Lemma 4.5 there exists an edge e' in \mathcal{D} with $H_1(\text{int}(e)) \subset \text{int}(e')$. Since the endpoints of $H_1(e)$ lie in \mathbf{V} , this implies that $e' = H_1(e)$.

To show that $e' = e$ we argue by contradiction and assume that $e \neq e'$. Then e and e' have the same endpoints, but no other points in common. Hence $\alpha = e \cup e'$ is a Jordan curve that contains two vertices, namely the endpoints of e and e' , but no other vertices. Let Ω_1 and Ω_2 be the two open Jordan regions that form the complementary components of α . Then both regions Ω_1 and Ω_2 contain vertices.

To see this note that the interior of every tile X is a connected set disjoint from the 1-skeleton E , and hence also disjoint from α . Hence $\text{int}(X)$ is contained in Ω_1 or Ω_2 . Moreover, since the union of the interiors of tiles is dense in S^2 , both regions Ω_1 and Ω_2 must contain the interior of at least one tile.

Now consider Ω_1 , for example, and pick a tile X with $\text{int}(X) \subset \Omega_1$. Then by our hypotheses the set $X \subset \overline{\Omega}_1 = \Omega_1 \cup \alpha$ contains at least three vertices. Since only two of them can lie on α , the set Ω_1 must contain a vertex. Similarly, Ω_2 must contain at least one vertex.

A contradiction can now be obtained from the fact that during the isotopy H the set $\text{int}(e)$ remains disjoint from the set of vertices, but on the other hand it has to sweep out one of the domains Ω_1 or Ω_2 and hence it meets a vertex.

To make this rigorous, we apply the topological fact mentioned before the statement of the lemma. Let D be the quotient of the product space $e \times I$ obtained by identifying all points (u, t) , $t \in I$, and by identifying all points (v, t) , $t \in I$, where u and v are the two endpoints of e . Then D is a two-dimensional cell. Since the isotopy H does not move the points u and v , the map $(p, t) \mapsto H_t(p)$ on $e \times I$ induces a continuous

map $\varphi: D \rightarrow S^2$. Moreover, $\varphi|_{\partial D}$ is a homeomorphism of $\partial \mathbb{D}$ onto α . Hence Ω_1 or Ω_2 is contained in the set

$$\varphi(\text{int}(D)) = \bigcup_{t \in (0,1)} H_t(\text{int}(e)).$$

In particular, the set $\varphi(\text{int}(D))$ contains a vertex. This is a contradiction, because we know that no set $H_t(\text{int}(e))$, $t \in I$, meets \mathbf{V} . Thus $H_1(e) = e$ as desired.

Having verified the statement about edges, it is now easy to see that $J = K$. Indeed, J is a union of edges in \mathcal{D} ; to see this consider the components of the set $J \setminus \mathbf{V}$. If γ is such component, then $\overline{\gamma} \setminus \gamma \subset \mathbf{V}$. Moreover, γ is contained in the 1-skeleton E , and does not meet the 0-skeleton \mathbf{V} . Again by Lemma 4.5 the set γ must be contained in the interior $\text{int}(e)$ of some edge e . This is only possible if $\gamma = \text{int}(e)$. Hence $\overline{\gamma} = e$. Since J is the union of the closures of these components γ , it follows that J is the union of edges e . For each such edge e we have $H_1(e) \subset K \subset E$ and so $H_1(e) = e$ by the first part of the proof. This implies $J \subset K$. Since J and K are Jordan curves, the desired identity $J = K$ follows. \square

11. THURSTON MAPS WITH INVARIANT CURVES

Let $f: S^2 \rightarrow S^2$ be a Thurston map, $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$, and \mathcal{D}^n be the cell decomposition of S^2 given by the n -cells for (f, \mathcal{C}) .

A set $M \subset S^2$ is called *f-invariant* (or simply *invariant* if f is understood) if

$$(11.1) \quad f(M) \subset M \quad \text{or equivalently} \quad M \subset f^{-1}(M).$$

Since the set $\text{post}(f)$ is f -invariant, we have

$$(11.2) \quad \text{post}(f) \subset f^{-1}(\text{post}(f)) \subset f^{-2}(\text{post}(f)) \subset \dots$$

We know (see Proposition 6.1 (iii)) that (11.2) is equivalent to the inclusions

$$\mathbf{V}^0 \subset \mathbf{V}^1 \subset \mathbf{V}^2 \subset \dots$$

for the vertex sets of the cell decompositions \mathcal{D}^n .

In general, a similar inclusion chain will not hold for the 1-skeleta $E^n := f^{-n}(\mathcal{C})$ of \mathcal{D}^n , but if \mathcal{C} is f -invariant, then we have

$$\mathcal{C} = E^0 \subset E^1 \subset E^2 \subset \dots$$

Actually, more is true as the following statement shows.

Proposition 11.1. *Let $k, n \in \mathbb{N}_0$, $f: S^2 \rightarrow S^2$ be a Thurston map, and $\mathcal{C} \subset S^2$ be an f -invariant Jordan curve with $\text{post}(f) \subset \mathcal{C}$. Then we have:*

- (i) $(\mathcal{D}^{n+k}, \mathcal{D}^k)$ is a cellular Markov partition for f^n .
- (ii) Every $(n+k)$ -tile X^{n+k} is contained in a unique k -tile X^k .
- (iii) Every k -tile X^k is equal to the union of all $(n+k)$ -tiles X^{n+k} with $X^{n+k} \subset X^k$.
- (iv) Every k -edge e^k is equal to the union of all $(n+k)$ -edges e^{n+k} with $e^{n+k} \subset e^k$.

Proof. (i) We know that the map f^n is cellular for $(\mathcal{D}^{n+k}, \mathcal{D}^n)$ (Proposition 6.1); so we have to show that \mathcal{D}^{n+k} is a refinement of \mathcal{D}^n (see Definition 4.6). By the invariance of \mathcal{C} we have $E^{n+k} = f^{-(n+k)}(\mathcal{C}) \supset E^k = f^{-k}(\mathcal{C})$, and so $S^2 \setminus E^{n+k} \subset S^2 \setminus E^k$.

To establish the first property of a refinement, we will show that every $(n+k)$ -cell is contained in some k -tile.

Let σ be an arbitrary $(n+k)$ -cell. If σ is a $(n+k)$ -tile, then $\text{int}(\sigma)$ is a connected set in $S^2 \setminus E^{n+k} \subset S^2 \setminus E^k$ and hence contained in the interior of a k -tile τ (see Proposition 6.1). It follows that $\sigma = \overline{\text{int}(\sigma)} \subset \tau$.

If σ is an $(n+k)$ -edge or an $(n+k)$ -vertex, then it is contained in an $(n+k)$ -tile (Lemma 5.1 (iv) and (v)), and hence in some k -tile by what we have just seen.

To establish the second property of a refinement, let τ be an arbitrary k -cell. We have to show that the $(n+k)$ -cells σ contained in τ cover τ .

If τ consists of a k -vertex p , then p is also an $(n+k)$ -vertex, and the statement is trivial.

If τ is a k -edge, consider the points in \mathbf{V}^{n+k} that lie on τ . Note that this includes the elements of $\partial\tau \subset \mathbf{V}^k \subset \mathbf{V}^{n+k}$. By using these points to partition τ , we can find finitely many arcs $\alpha_1, \dots, \alpha_N$ such that $\tau = \alpha_1 \cup \dots \cup \alpha_N$, each arc α_i has endpoints in $\mathbf{V}^{n+k} \supset \mathbf{V}^k$ and has interior $\text{int}(\alpha_i)$ disjoint from \mathbf{V}^{n+k} . Then for each $i = 1, \dots, N$ the set $\text{int}(\alpha_i)$ is a connected set in $E^k \setminus \mathbf{V}^{n+k} \subset E^{n+k} \setminus \mathbf{V}^{n+k}$. It follows that $\text{int}(\alpha_i)$ and hence also α_i is contained in some $(n+k)$ -edge σ_i (Proposition 6.1 (v)). Since the endpoints of α_i lie in \mathbf{V}^k , they cannot lie in $\text{int}(\sigma_i)$, and so they are also endpoints of σ_i . This implies that $\alpha_i = \sigma_i$. In particular, the $(n+k)$ -edges $\sigma_1, \dots, \sigma_N$ are contained in τ and form a cover of τ . The statement follows in this case.

Finally, let τ be a k -tile. If $p \in \text{int}(\tau)$ is arbitrary, then p is contained in an $(n+k)$ -tile σ . By the first part of the proof, σ is contained in a k -cell. Since τ is the only k -cell that contains p , we must have $\sigma \subset \tau$. This implies that the union of the $(n+k)$ -cells contained in τ cover

$\text{int}(\tau)$. On the other hand, this union consists of finitely many tiles and is hence a closed set. It follows that the union also contains $\overline{\text{int}(\tau)} = \tau$.

(ii) We have just seen that every $(n+k)$ -tile X^{n+k} is contained in a k -tile X^k . This tile is unique. For suppose \tilde{X}^k is another k -tile with $X^{n+k} \subset \tilde{X}^k$. Then

$$\emptyset \neq \text{int}(X^{n+k}) \subset \text{int}(X^k) \cap \text{int}(\tilde{X}^k),$$

and so X^k and \tilde{X}^k have common interior points. This implies $X^k = \tilde{X}^k$.

(iii)–(iv): Both statements were established in the proof of (i). \square

Let f and \mathcal{C} be as in the previous proposition. Then by (i) the pair $(\mathcal{D}^1, \mathcal{D}^0)$ is a cellular Markov partition for f , and this partition generates the cell decompositions \mathcal{D}^n as in Proposition 4.10. If X^n is any n -tile, then by (ii) there exist unique i -tiles X^i for $i = 0, \dots, n-1$ such that

$$X^n \subset X^{n-1} \subset \dots \subset X^0.$$

We refer to the statements (iii) and (iv) informally by saying that tiles and edges are “subdivided” by tiles and edges of higher order. Let $\mathcal{S} = \mathcal{S}(f, \mathcal{C})$ denote the set of all sequences $\{X^n\}$, where X^n is an n -tile for $n \in \mathbb{N}_0$ and

$$X^0 \supset X^1 \supset X^2 \supset \dots$$

Since tiles are subdivided by tiles of higher order, for each point $p \in S^2$ we can find a sequence $\{X^n\} \in \mathcal{S}$ such that $p \in \bigcap_n X^n$. Here it is understood that the intersection is taken over all $n \in \mathbb{N}_0$. In the following we use a similar convention for intersections of sets labeled by some index n , k , etc., if the range of the indices is clear from the context.

In general, a sequence $\{X^n\} \in \mathcal{S}$ that contains a given point $p \in S^2$ is not unique. Moreover, the intersection $\bigcap_n X^n$ may contain more than one point. It turns out that this gives a criterion when f is expanding.

Lemma 11.2. *Let $f: S^2 \rightarrow S^2$ be a Thurston map, and $\mathcal{C} \subset S^2$ be an f -invariant Jordan curve with $\text{post}(f) \subset \mathcal{C}$. Then the map f is expanding if and only if for each sequence $\{X^n\} \in \mathcal{S}(f, \mathcal{C})$ the intersection $\bigcap_n X^n$ consists of precisely one point.*

Proof. Fix a metric on S^2 that induces the standard topology on S^2 . If f is expanding and $\{X^n\} \in \mathcal{S} := \mathcal{S}(f, \mathcal{C})$, then $\text{diam}(X^n) \rightarrow 0$ as $n \rightarrow \infty$. Hence $\bigcap_n X^n$ cannot contain more than one point. On the other hand, this set is an intersection of a nested sequence of nonempty

compact sets and hence nonempty. So the set $\bigcap_n X^n$ contains precisely one point.

For the converse direction suppose that $\bigcap_n X^n$ is a singleton set for each sequence $\{X^n\} \in \mathcal{S}$. To establish that f is expanding we have to show that

$$\lim_{n \rightarrow \infty} \max \{\text{diam}(X) : X \text{ is an } n\text{-tile}\} = 0.$$

We argue by contradiction and assume that this is not the case. Then there exists $\delta > 0$ such that $\text{diam}(X) \geq \delta$ for some tiles X of arbitrarily high order.

We define a descending sequence of tiles $X^0 \supset X^1 \supset X^2 \supset \dots$ as follows. Let X^0 be a 0-tile such that X^0 contains tiles X of arbitrarily high order with $\text{diam}(X) \geq \delta$. Since the tiles are subdivided by tiles of higher order, and so every tile is contained in one of the finitely many 0-tiles (in our case there actually two 0-tiles), there exists such a 0-tile. Note that then $\text{diam}(X^0) \geq \delta$. Moreover, among the finitely many 1-tiles into which X^0 is subdivided there must be a 1-tile $X^1 \subset X^0$ such that X^1 contains tiles X of arbitrarily high order with $\text{diam}(X) \geq \delta$. Again this implies that $\text{diam}(X^1) \geq \delta$. Repeating this procedure we obtain a sequence $\{X^n\} \in \mathcal{S}$ such that $\text{diam}(X^n) \geq \delta$ for all $n \in \mathbb{N}_0$. It is easy to see that this implies that the set $\bigcap_n X^n$ also has diameter $\geq \delta > 0$, and so it contains at least two points. This is a contradiction showing that f is expanding. \square

Let $f: S^2 \rightarrow S^2$ be a Thurston map, $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset S^2$, and assume that $\#\text{post}(f) \geq 3$. Recall the definition of the numbers $D_n = D_n(f, \mathcal{C})$ in (7.4). We know that $D_n \rightarrow \infty$ if f is expanding (see Lemma 8.3). If f is not necessarily expanding, but the Jordan curve \mathcal{C} used in the definition of D_n is invariant, then it follows from the previous discussion that the numbers D_n are increasing, i.e., $D_{n+1} \geq D_n$ for all $n \in \mathbb{N}_0$. Moreover, one can show exponential increase of the numbers D_n under the additional assumption that there exists $n_0 \in \mathbb{N}$ with $D_{n_0} \geq 2$. This is the content of the following lemma.

Lemma 11.3. *Let $f: S^2 \rightarrow S^2$ be a Thurston map with $\#\text{post}(f) \geq 3$, let $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$ and suppose that \mathcal{C} is f -invariant, and let $D_n = D_n(f, \mathcal{C})$ for $n \in \mathbb{N}_0$. Then for all $n, k \in \mathbb{N}_0$ we have*

$$D_{n+k} \geq D_n D_k$$

if $\#\text{post}(f) \geq 4$, and

$$(11.3) \quad D_{n+k} \geq D_n(D_k - 1) + 1$$

if $\#\text{post}(f) = 3$.

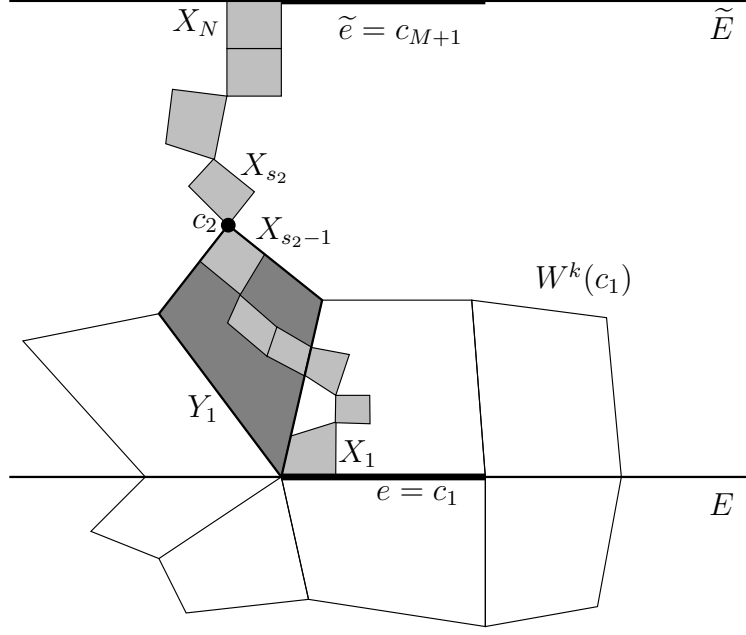


FIGURE 4. The proof of Lemma 11.3.

Moreover,

$$\alpha := \lim_{n \rightarrow \infty} \frac{1}{n} \log(D_n) = \sup_{n \in \mathbb{N}} \frac{1}{n} \log(D_n) \leq \log(\deg(f)).$$

If in addition there exists $n_0 \in \mathbb{N}$ with $D_{n_0} \geq 2$, then $\alpha > 0$ and $D_n \rightarrow \infty$ as $n \rightarrow \infty$.

Before we prove this lemma let us fix some terminology. An n -chain is a finite sequence of n -tiles X_1, \dots, X_N , where $X_i \cap X_{i+1} \neq \emptyset$ for $i = 1, \dots, N-1$. It joins two disjoint sets A and B if $A \cap X_1 \neq \emptyset$ and $B \cap X_N \neq \emptyset$. The chain joins two points x and y if it joins the sets $\{x\}$ and $\{y\}$. The n -chain is called a *simple chain* joining A and B if there is no proper subsequence of X_1, \dots, X_N that is also a chain joining A and B . If we put $X_{-1} := A$ and $X_{N+1} := B$, then this is equivalent with the requirement that $X_i \cap X_j = \emptyset$ whenever $-1 \leq i < j \leq N+1$ and $j - i \geq 2$.

Proof of Lemma 11.3. Case 1. $\# \text{post}(f) \geq 4$.

Let X_1, \dots, X_N be a set of $(n+k)$ -tiles whose union is connected and joins opposite sides of \mathcal{C} . We may assume that these tiles form a chain joining disjoint 0-edges E and \tilde{E} . To prove the desired inequality we will break this chain into M subchains $X_{s_i}, \dots, X_{s_{i+1}-1}$, where $M \in \mathbb{N}$, $i = 1, \dots, M$, and $s_1 = 1 < s_2 < \dots < s_{M+1} = N+1$. The length

of each subchain (i.e., the number $s_{i+1} - s_i$) will be at least D_n . The number M of subchains will be at least D_k . Thus $N \geq D_n D_k$, and since the minimum over all N is equal to D_{n+k} , the desired inequality follows.

To achieve the desired bound on the length we will ensure that each subchain $X_{s_i}, \dots, X_{s_{i+1}-1}$ joins disjoint k -cells. Then the length of such a subchain is at least D_n by Lemma 7.10.

To control the number of subchains, we will associate to each one a k -tile Y_i . These k -tiles Y_1, \dots, Y_M will form a k -chain joining E and \tilde{E} , and hence opposite sides of \mathcal{C} . Thus M , which is the number of k -tiles in this chain, as well as the number of subchains, is at least D_k (by definition of this quantity; see (7.4)).

We now provide the details of the construction, which is illustrated in Figure 4. We will use auxiliary k -cells c_1, c_2, \dots of dimension ≤ 1 . If c_i is 0-dimensional, then c_i consists of a k -vertex p_i , and we let $W^k(c_i) := W^k(p_i)$ (see Definition 7.1). If c_i is 1-dimensional, then c_i is a k -edge and $W^k(c_i)$ is the edge flower of c_i as in Definition 7.4.

Since \mathcal{C} is f -invariant, the cell decomposition \mathcal{D}^k is a refinement of \mathcal{D}^0 . Hence there exist disjoint k -edges $e \subset E$ and $\tilde{e} \subset \tilde{E}$ with $X_1 \cap e \neq \emptyset$ and $X_N \cap \tilde{e} \neq \emptyset$.

For some number $M \in \mathbb{N}$ we will now inductively define k -cells c_1, \dots, c_{M+1} of dimension ≤ 1 , k -tiles Y_1, \dots, Y_M , and indices $s_1 = 1 < s_2 < \dots < s_{M+1} = N + 1$ with the following properties:

- (i) $c_1 = e$, $c_{M+1} = \tilde{e}$, and $c_i \cap c_{i+1} = \emptyset$ for $i = 1, \dots, M$.
- (ii) $c_i \cap Y_i \neq \emptyset$ for $i = 1, \dots, M$, $c_{i+1} \subset \partial Y_i$ for $i = 1, \dots, M - 1$, and $\tilde{e} \cap Y_M \neq \emptyset$.
- (iii) $X_{s_i}, \dots, X_{s_{i+1}-1}$ is an $(n + k)$ -chain joining c_i and c_{i+1} for $i = 1, \dots, M$.

Note that (i) and (ii) imply that $E \cap Y_1 \supset e \cap Y_1 \neq \emptyset$, $\tilde{E} \cap Y_M \supset \tilde{e} \cap Y_M \neq \emptyset$, and $Y_i \cap Y_{i+1} \supset c_{i+1} \cap Y_{i+1} \neq \emptyset$ for $i = 1, \dots, M - 1$. Hence Y_1, \dots, Y_M will be a k -chain joining the 0-edges E and \tilde{E} as desired.

Let $s_1 = 1$ and $c_1 = e$. Suppose first that \tilde{e} meets $\overline{W^k(c_1)}$. Since \tilde{e} is disjoint from $e = c_1$ and hence from $W^k(c_1)$, the points in $\tilde{e} \cap \overline{W^k(c_1)}$ lie in $\partial W^k(c_1)$. By Lemma 7.5 (ii) there exists a k -tile Y_1 that meets both c_1 and $\tilde{e} \supset \tilde{e} \cap \overline{W^k(c_1)}$. We let $M = 1$, set $c_2 = \tilde{e}$, and stop the construction. We have all the desired properties (i)–(iii).

In the other case where $\tilde{e} \cap \overline{W^k(c_1)} = \emptyset$ not all the $(n + k)$ -tiles X_1, \dots, X_N are contained in $\overline{W^k(c_1)}$. So there exists a smallest index $s_2 \geq 1$ such that X_{s_2} meets $S^2 \setminus \overline{W^k(c_1)}$. Then $s_2 > s_1 = 1$, because

X_1 meets $e = c_1$ and is hence contained in $\overline{W^k(c_1)}$. To see this we use Lemma 7.5 (iii) and the fact that X_1 is contained in some k -tile. Moreover, for a similar reason we have $X_{s_2} \subset S^2 \setminus W^k(c_1)$. By definition of s_2 the set X_{s_2-1} is contained in $\overline{W^k(c_1)}$. Hence every point in the nonempty intersection $X_{s_2-1} \cap X_{s_2}$ lies in $\partial W^k(c_1)$. In particular, by Lemma 7.5 (iii) there exists a k -cell $c_2 \subset \partial W^k(c_1)$ of dimension ≤ 1 that has common points with both X_{s_2-1} and X_{s_2} , and a k -tile $Y_1 \subset \overline{W^k(c_1)}$ with $c_1 \cap Y_1 \neq \emptyset$ and $c_2 \subset \partial Y_1$. Then $c_1 \cap c_2 = \emptyset = c_2 \cap \tilde{e}$, and the chain $X_{s_1} = X_1, \dots, X_{s_2-1}$ joins c_1 and c_2 .

We can now repeat the construction as in the first step by using the chain X_{s_2}, \dots, X_N that joins the disjoint k -cells c_2 and \tilde{e} , etc. If in the process one of the cells c_i has dimension 0, we invoke Lemma 7.2 (ii) and (iii) instead of Lemma 7.5 (ii) and (iii) in the above construction. The construction eventually stops, and it is clear that we obtain cells and indices with the desired properties.

Case 2. $\# \text{post}(F) = 3$.

Let E_1, E_2, E_3 be the three 0-edges. Consider a connected union K of $(n+k)$ -tiles joining opposite sides of \mathcal{C} with $N = D_{n+k}$ elements. Then K meets k -edges contained in the 0-edges, say k -edges $e_i \subset E_i$ for $i = 1, 2, 3$. From K we can extract a simple $(n+k)$ -chain joining e_1 and e_2 as well as another simple chain that joins e_3 to one tile X in the chain joining e_1 and e_2 . Starting from this “center tile” X , we can find three simple $(n+k)$ -chains that join X to the edges e_1, e_2, e_3 , respectively, and have only the tile X in common.

More precisely, for $i = 1, 2, 3$ we can find $N_i \in \mathbb{N}_0$ and $(n+k)$ -chains $X, X_1^i, \dots, X_{N_i}^i$ that join X and e_i . Here the first tile X is the same in all chains and it is understood that the chain consists only of X if $N_i = 0$. Moreover, all the $(n+k)$ -tiles

$$X, X_1^1, \dots, X_{N_1}^1, X_1^2, \dots, X_{N_2}^2, X_1^3, \dots, X_{N_3}^3$$

are pairwise distinct tiles from K . Thus their number is bounded by the number of $(n+k)$ -tiles in K . Since they still form a connected set joining opposite sides of \mathcal{C} , we have $N_1 + N_2 + N_3 + 1 = N = D_{n+k}$.

Pick a k -tile Y with $X \subset Y$, and consider the chain $X, X_1^1, \dots, X_{N_1}^1$. Suppose that $Y \cap e_1 = \emptyset$. Since $X \subset Y$ we have $N_1 \geq 1$ and the chain X_1, \dots, X_{N_1} joins Y and e_1 . Hence this chain or a subchain must also join a k -edge $e \subset \partial Y$ and e_1 . Then $e \cap e_1 = \emptyset$. As in the first part of the proof we can find k -tiles Y_1, \dots, Y_{M_1} joining e and e_1 , where $M_1 \in \mathbb{N}$ and $N_1 \geq M_1 D_n$.

If $Y \cap e_1 \neq \emptyset$, we set $M_1 = 0$ and do not define new k -tiles. In any case we have that $Y, Y_1^1, \dots, Y_{M_1}^1$ is a chain joining Y and e_1 (again we

use the convention that this chain consists only of Y if $M_1 = 0$). We also have $N_1 \geq M_1 D_n$ (which is trivial if $M_1 = 0$).

Using a similar construction for the other indices $i = 2, 3$, we obtain numbers $M_i \in \mathbb{N}_0$ for each $i = 1, 2, 3$ that satisfy $N_i \geq M_i D_n$, and chains $Y, Y_1^i, \dots, Y_{M_i}^i$ of k -tiles that join Y and e_i . The union of these k -tiles is a connected set joining opposite sides of \mathcal{C} . Hence it contains at least D_k distinct elements. On the other hand, the number of distinct k -tiles in the union is at most $M_1 + M_2 + M_3 + 1$ (note that the three chains may have other k -tiles in common apart from Y). Hence $D_k \leq M_1 + M_2 + M_3 + 1$, and it follows that

$$D_n(D_k - 1) + 1 \leq D_n(M_1 + M_2 + M_3) + 1 \leq N_1 + N_2 + N_3 + 1 \leq D_{n+k},$$

which is the desired inequality (11.3).

In order to prove the remaining statements first note that inequality (11.3) is also true if $\text{post}(f) = 4$. A simple induction argument using (11.3) shows that if $D_N \geq 2$ for some $N \in \mathbb{N}$, then

$$(11.4) \quad D_{kN} \geq D_N^{k-1} + 1$$

for all $k \in \mathbb{N}$. For such N let $k(n) = \lfloor n/N \rfloor$. Noting that the sequence $\{D_n\}$ is non-decreasing and using (11.4), we obtain

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log(D_n) &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log(D_{k(n)N}) \\ &\geq \liminf_{n \rightarrow \infty} \frac{k(n) - 1}{n} \log(D_N) = \frac{1}{N} \log(D_N). \end{aligned}$$

This inequality is trivially true if $D_N = 1$, and so

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log D_n \geq \sup_{n \in \mathbb{N}} \frac{1}{n} \log D_n.$$

On the other hand,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log(D_n) \leq \sup_{n \in \mathbb{N}} \frac{1}{n} \log(D_n),$$

and so

$$\alpha = \lim_{n \rightarrow \infty} \frac{1}{n} \log(D_n) = \sup_{n \in \mathbb{N}} \frac{1}{n} \log(D_n).$$

Note that $D_n \leq \#\mathbf{X}^n \leq 2 \deg(f)^n$ which implies $\alpha \leq \log(\deg(f))$.

Finally, if there exists $n_0 \in \mathbb{N}$ with $D_{n_0} \geq 2$, then

$$\alpha = \sup_{n \in \mathbb{N}} \frac{1}{n} \log(D_n) \geq \frac{1}{n_0} \log(D_{n_0}) > 0,$$

and it is clear from the definition of α that $D_n \rightarrow \infty$ as $n \rightarrow \infty$. \square

The last lemma shows that if f is a Thurston map with an invariant Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$, and there exists $n_0 \in \mathbb{N}$ such that $D_{n_0} = D_{n_0}(f, \mathcal{C}) \geq 2$, then the numbers $D_n = D_n(f, \mathcal{C})$ are actually exponentially increasing; indeed, if $\alpha > 0$ is as in the lemma and $\epsilon > 0$ is arbitrary, then $D_n \gtrsim e^{(\alpha-\epsilon)n}$ for large n .

This situation will be important enough to warrant a separate definition.

Definition 11.4 (Combinatorial expansion). Let $f: S^2 \rightarrow S^2$ be a Thurston map. We call f *combinatorially expanding* if $\text{post}(f) \geq 3$, and there exists a Jordan curve $\mathcal{C} \subset S^2$ that is f -invariant, satisfies $\text{post}(f) \subset \mathcal{C}$, and for which there is a number $n_0 \in \mathbb{N}$ such that $D_{n_0}(f, \mathcal{C}) \geq 2$.

If f and \mathcal{C} are as in the previous definition, then we say that f is *combinatorially expanding for \mathcal{C}* .

By definition of $D_n(f, \mathcal{C})$ the condition $D_{n_0}(f, \mathcal{C}) \geq 2$ means that no single n_0 -tile joins opposite sides of \mathcal{C} . If a map f as in Definition 11.4 is expanding and $\mathcal{C} \subset S^2$ is an f -invariant Jordan curve with $\text{post}(f) \subset \mathcal{C}$, then f is also combinatorially expanding for \mathcal{C} (in this case $D_n(f, \mathcal{C}) \rightarrow \infty$; see Lemma 8.3). The converse is not true in general, since a combinatorially expanding Thurston map need not be expanding. However, we will see in Section 13 that each combinatorially expanding Thurston is equivalent to an expanding Thurston map with an invariant curve.

The condition of combinatorial expansion is indeed combinatorial in nature, because it can be verified just by knowing the combinatorics of the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$, $n \in \mathbb{N}_0$. This in turn is determined by the combinatorics of the pair $(\mathcal{D}^1, \mathcal{D}^0)$ and the map $c \in \mathcal{D}^1 \mapsto f(c) \in \mathcal{D}^0$ (see Remark 4.11).

We finish this section with a lemma on combinatorially expanding Thurston maps that will be useful later.

Lemma 11.5. *Let $f: S^2 \rightarrow S^2$ be a Thurston map that is combinatorially expanding for the Jordan curve $\mathcal{C} \subset S^2$, and suppose that $D_{n_0}(f, \mathcal{C}) \geq 2$ where $n_0 \in \mathbb{N}$.*

- (i) *If $n \in \mathbb{N}_0$ and e is an n -edge, then there exists an $(n+n_0)$ -vertex p with $p \in \text{int}(e)$.*
- (ii) *If $n \in \mathbb{N}_0$ and X is an n -tile, then there exists an $(n+n_0)$ -edge with $\text{int}(e) \subset \text{int}(X)$, and an $(n+2n_0)$ -vertex p with $p \in \text{int}(X)$.*

Proof. In the previous statements and the ensuing proof it is understood that the term k -cell for $k \in \mathbb{N}_0$ refers to a cell in $\mathcal{D}^k = \mathcal{D}^k(f, \mathcal{C})$.

(i) Suppose e is an n -edge that does not contain $(n + n_0)$ -vertices in its interior. By Proposition 11.1 (iv) the n -edge e is equal to the union of all $(n + n_0)$ -edges contained in e . Thus e must be an $(n + n_0)$ -edge itself. Let u and v be the endpoints of e , and X be an $(n + n_0)$ -tile containing e in its boundary. Then $K = X$ meets the two disjoint n -cells $\{u\}$ and $\{v\}$. Hence by Lemma 7.10 the set K should consist of at least $D_{n_0} = D_{n_0}(f, \mathcal{C}) \geq 2$ $(n + n_0)$ -tiles. This is a contradiction proving the statement.

(ii) Let X be an n -tile. By Proposition 11.1 (iii) we know that X is the union of all $(n + n_0)$ -tiles contained in X . In particular, there exists an $(n + n_0)$ -tile Y with $Y \subset X$. We claim that there exists an $(n + n_0)$ -edge in the boundary of Y that meets $\text{int}(X)$. Otherwise, $\partial Y \cap \text{int}(X) = \emptyset$, and as $Y \subset X$, we must have $\partial Y \subset \partial X$. Since both sets ∂Y and ∂X are Jordan curves, this is only possible if $\partial Y = \partial X$. Then Y meets all n -vertices contained in ∂X , and two distinct n -vertices in particular. As in the proof of (i), this leads to a contradiction.

Hence there exists an $(n + n_0)$ -edge e with $e \cap \text{int}(X) \neq \emptyset$. Since $\text{int}(X)$ is an open subset of S^2 , we then also have $\text{int}(e) \cap \text{int}(X) \neq \emptyset$. Since \mathcal{D}^{n+n_0} is a refinement of \mathcal{D} , by Lemma 4.7 we know that there is a unique cell τ in \mathcal{D}^n with $\text{int}(e) \subset \text{int}(\tau)$. Then $\text{int}(\tau) \cap \text{int}(X) \neq \emptyset$, and so $X = \tau$ by Lemma 4.2. Hence $\text{int}(e) \subset \text{int}(X)$ as desired.

By (i) there exists an $(n + 2n_0)$ -vertex p with $p \in \text{int}(e)$. Then we also have $p \in \text{int}(X)$ as desired. \square

12. TWO-TILE SUBDIVISION RULES

We have seen how a Thurston map f and a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$ can be used to define cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ of S^2 . If \mathcal{C} is f -invariant, then $(\mathcal{D}^1, \mathcal{D}^0)$ is a cellular Markov partition for f .

In this section we will see that this process can be reversed. Starting with a pair $(\mathcal{D}^1, \mathcal{D}^0)$ of cell decompositions of S^2 , one can show that under suitable additional assumptions there exists a postcritically-finite branched covering map that is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$. As before we will refer to the elements in \mathcal{D}^0 as the 0-cells and the elements in \mathcal{D}^1 as the 1-cells, and speak of 1-edges, 0-tiles, etc. The map f is unique up to Thurston equivalence if additional data is provided, namely a *labeling* $\mathcal{D}^1 \rightarrow \mathcal{D}^0$. The concept of a labeling extracts the relevant combinatorial properties of the map $\tau \in \mathcal{D}^1 \mapsto f(\tau) \in \mathcal{D}^0$ if f is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$. Here is the precise general definition.

Definition 12.1 (Labelings). Let \mathcal{D}^1 and \mathcal{D}^0 be cell complexes. Then a *labeling* of $(\mathcal{D}^1, \mathcal{D}^0)$ is map $L: \mathcal{D}^1 \rightarrow \mathcal{D}^0$ satisfying the following conditions:

- (i) $\dim(L(\tau)) = \dim(\tau)$ for all $\tau \in \mathcal{D}^1$,
- (ii) if $\sigma, \tau \in \mathcal{D}^1$ and $\sigma \subset \tau$, then $L(\sigma) \subset L(\tau)$.
- (iii) if $\sigma, \tau, c \in \mathcal{D}^1$, $\sigma, \tau \subset c$ and $L(\sigma) = L(\tau)$, then $\sigma = \tau$.

So a labeling is a map $L: \mathcal{D}^1 \rightarrow \mathcal{D}^0$ that preserves inclusions and dimensions of cells, and is “injective on cells” $c \in \mathcal{D}^1$ in the sense of (iii). In particular, every cell of dimension 0 in \mathcal{D}^1 is mapped to a cell of dimension 0 in \mathcal{D}^0 . If v is a *vertex* in \mathcal{D}^1 , i.e., if $\{v\}$ is a cell of dimension 0 in \mathcal{D}^1 , then we can write $L(\{v\}) = \{w\}$, where w is a vertex in \mathcal{D}^0 . We define $L(v) = w$. In the following we always assume that a labeling $L: \mathcal{D}^1 \rightarrow \mathcal{D}^0$ has been extended to the set of vertices of \mathcal{D}^1 in this way; this will allow us to ignore the subtle distinction between vertices and cells of dimension 0, i.e., sets consisting of one vertex.

If we have a labeling $L: \mathcal{D}^1 \rightarrow \mathcal{D}^0$ we should think of each element $\tau \in \mathcal{D}^1$ as “carrying” the label $L(\tau) \in \mathcal{D}^0$. In applications it is often more intuitive and convenient to allow more general index sets \mathcal{L} of the same cardinality \mathcal{D}^0 as labeling sets for the elements in \mathcal{D}^1 . In such situations we fix a bijection $\psi: \mathcal{D}^0 \rightarrow \mathcal{L}$ and call a map $L': \mathcal{D}^1 \rightarrow \mathcal{L}$ a labeling if $\psi^{-1} \circ L': \mathcal{D}^1 \rightarrow \mathcal{D}^0$ is a labeling in the sense of Definition 12.1. We will discuss this further below.

Related to labelings is the concept of an isomorphism between cell complexes.

Definition 12.2 (Isomorphisms of cell complexes). Let \mathcal{D} and \mathcal{D}' be cell complexes. A bijection $\phi: \mathcal{D} \rightarrow \mathcal{D}'$ is called an *isomorphism (of cell complexes)* if the following conditions are satisfied:

- (i) $\dim(\phi(\tau)) = \dim(\tau)$ for all $\tau \in \mathcal{D}$,
- (ii) if $\sigma, \tau \in \mathcal{D}$, then $\sigma \subset \tau$ if and only if $\phi(\sigma) \subset \phi(\tau)$.

If we are given another cell complex \mathcal{D}^0 and labelings $L: \mathcal{D} \rightarrow \mathcal{D}^0$ and $L': \mathcal{D}' \rightarrow \mathcal{D}^0$, then an isomorphism $\phi: \mathcal{D} \rightarrow \mathcal{D}'$ is called *label-preserving* if $L = L' \circ \phi$.

Let S^2 be an oriented 2-sphere, and \mathcal{D} be a cell decomposition of S^2 . Recall (see Section 5) that a flag in \mathcal{D} is a triple (c_0, c_1, c_2) , where c_i is a cell in \mathcal{D} of dimension i for $i = 0, 1, 2$ and $c_0 \subset c_1 \subset c_2$. In this case $c_0 = \{v\}$ consists of a vertex v of \mathcal{D} which must be one of the endpoints of c_1 , and c_1 is oriented by considering v as the initial and the other vertex in ∂c_1 as the terminal point of c_1 . The cell c_2 is

one of the two tiles in \mathcal{D} that contain c_1 in their boundary. The flag (c_0, c_1, c_2) is positively-oriented if c_2 lies on the left of the oriented edge c_1 according to the given orientation of S^2 .

If $L: \mathcal{D}^1 \rightarrow \mathcal{D}^0$ is a labeling of a pair $(\mathcal{D}^1, \mathcal{D}^0)$ of cell decompositions of S^2 and (c_0, c_1, c_2) is a flag in \mathcal{D}^1 , then $(L(c_0), L(c_1), L(c_2))$ is a flag in \mathcal{D}^0 . This follows from the definition of a labeling. So a labeling maps “flags to flags”. We say that the labeling is *orientation-preserving* if it maps positively-oriented flags in \mathcal{D}^1 to positively-oriented flags in \mathcal{D}^0 .

If $f: S^2 \rightarrow S^2$ is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$, then f induces a natural labeling $L: \mathcal{D}^1 \rightarrow \mathcal{D}^0$ given by $L(\tau) = f(\tau)$ for $\tau \in \mathcal{D}^1$. Moreover, if in addition $f|_X$ is orientation-preserving for each tile X in \mathcal{D}^1 (which is always true if f is a branched covering map), then this labeling L is orientation-preserving.

If a labeling $L: \mathcal{D}^1 \rightarrow \mathcal{D}^0$ is given, then we say that a map $f: S^2 \rightarrow S^2$ that is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$ is *compatible* with the labeling L if $L(\tau) = f(\tau)$ for each $\tau \in \mathcal{D}^1$, i.e., if the labeling induced by f is equal to the given labeling.

In the proof of the next proposition we need some simple facts about homeomorphisms and isotopies. If α is an arc, then every homeomorphism $\varphi: \alpha \rightarrow \alpha$ that fixes the endpoints of α is isotopic to the identity rel. $\partial\alpha$. Indeed, we may assume that α is equal to the unit interval $I = [0, 1]$. Then $\varphi(0) = 0$, $\varphi(1) = 1$, and φ is strictly increasing on $[0, 1]$. Define $H: I \times I \rightarrow I$ by

$$H(s, t) = (1 - t)\varphi(s) + ts$$

for $s, t \in I$. Then for each $t \in I$, the map $H_t = H(\cdot, t)$ is strictly increasing on I . It follows that H is an isotopy. We have $H_t(0) = 0$ and $H_t(1) = 1$ for all $t \in I$, and $H_0 = \varphi$ and $H_1 = \text{id}_I$. Hence φ and id_I are isotopic rel. $\partial I = \{0, 1\}$ by the isotopy H .

Let $X \subset S^2$ be a closed Jordan region. If $h: \partial X \times I \rightarrow \partial X$ is an isotopy with $h(\cdot, 0) = \text{id}_{\partial X}$, then there exists an isotopy $H: X \times I \rightarrow X$ such that $H(\cdot, 0) = \text{id}_X$ and $H(p, t) = h(p, t)$ for all $p \in \partial X$ and $t \in I$. So an isotopy h on the boundary of X with $h_0 = \text{id}_{\partial X}$ can be extended to an isotopy H on X with $H_0 = \text{id}_X$. To see this, we may assume that $X = \mathbb{D}$. Then H is obtained from h by radial extension; more precisely, we define

$$H(re^{is}, t) = rh(e^{is}, t)$$

for all $r \in [0, 1]$ and $s \in [0, 2\pi]$. Then H is well-defined and it is easy to see that H is an isotopy with the desired properties. By using the Schönflies theorem and a similar radial extension one can also show that if X and X' are closed Jordan regions in S^2 , then every homeomorphism $\varphi: \partial X \rightarrow \partial X'$ extends to a homeomorphism $\Phi: X \rightarrow X'$.

If $\varphi: X \rightarrow X$ is a homeomorphism with $\varphi|_{\partial X} = \text{id}_{\partial X}$, then φ is isotopic to id_X rel. ∂X . Indeed, again we may assume that $X = \overline{\mathbb{D}}$. Then we obtain the desired isotopy by the “Alexander trick”: for $z \in \overline{\mathbb{D}}$ and $t \in I$ we define $H(z, t) = t\varphi(z/t)$ if $|z| < t$, and $H(z, t) = z$ if $|z| \geq t$. It is easy to see that H is an isotopy rel. $\partial \mathbb{D}$ with $H_0 = \text{id}_X$ and $H_1 = \varphi$.

If $\varphi, \tilde{\varphi}: X \rightarrow X$ are two homeomorphisms with $\varphi|_{\partial X} = \tilde{\varphi}|_{\partial X}$, then we can apply the previous remark to $\psi = \tilde{\varphi} \circ \varphi^{-1}$ and conclude that φ and $\tilde{\varphi}$ are isotopic rel. ∂X .

Proposition 12.3. *Let $(\mathcal{D}^1, \mathcal{D}^0)$ be a pair of cell decompositions of an oriented 2-sphere S^2 with an orientation-preserving labeling. Assume that every vertex of \mathcal{D}^0 is also a vertex of \mathcal{D}^1 .*

Then there exists a postcritically-finite branched covering map $f: S^2 \rightarrow S^2$ that is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$ and is compatible with the given labeling. The map f is unique up to Thurston equivalence.

Note that f is a Thurston map if $\deg(f) \geq 2$, i.e., if f is not a homeomorphism.

Proof. The existence of a map f as desired follows from the well-known procedure of successive extensions on the skeleta of the cell decomposition \mathcal{D}^1 . Indeed, let $L: \mathcal{D}^1 \rightarrow \mathcal{D}^0$ be an orientation-preserving labeling. If $v \in S^2$ is a 1-vertex (i.e., a vertex in \mathcal{D}^1), then $L(v)$ is a 0-vertex (i.e., a vertex in \mathcal{D}^0). Set $f(v) = L(v)$. This defines f on the 0-skeleton of \mathcal{D}^1 . To extend this to the 1-skeleton of \mathcal{D}^1 , let e be an arbitrary 1-edge. Then $e' = L(e)$ is a 0-edge. Moreover, if u and v are the 1-vertices that are the endpoints of e , then $u' = f(u) = L(u)$ and $v' = f(v) = L(v)$ are distinct 0-vertices contained in e' . Hence they are the endpoints of e' . So we can extend f to e by choosing a homeomorphism of e onto e' that agrees with f on the endpoints of e . In this way we can continuously extend f to the 1-skeleton of \mathcal{D}^1 so that $f|_{\tau}$ is a homeomorphism of τ onto $L(\tau)$ whenever $\tau \in \mathcal{D}^1$ is a cell of dimension ≤ 1 .

If X is an arbitrary 1-tile, then ∂X is a subset of the 1-skeleton of \mathcal{D}^1 and hence f is already defined on ∂X . Then $f|_{\partial X}$ is a continuous mapping of ∂X into the boundary $\partial X'$ of the 0-tile $X' = L(X)$. The map $f|_{\partial X}$ is injective. Indeed, suppose that $u, v \in \partial X$ and $f(u) = f(v)$. Then there exist unique 1-cells $\sigma, \tau \subset \partial X$ of dimension ≤ 1 such that $u \in \text{int}(\sigma)$ and $v \in \text{int}(\tau)$. Then

$$f(u) = f(v) \in \text{int}(f(\sigma)) \cap \text{int}(f(\tau)) = \text{int}(L(\sigma)) \cap \text{int}(L(\tau))$$

and so the 1-cells $L(\sigma)$ and $L(\tau)$ must be the same. Since L is a labeling and $\sigma, \tau \subset X \in \mathcal{D}^1$, it follows that $\sigma = \tau$. As the map f restricted to the 1-cell $\sigma = \tau$ is injective, we conclude $u = v$ as desired.

Since every injective and continuous map of a Jordan curve into another Jordan curve is surjective, it follows that $f|_{\partial X}$ is a homeomorphism of ∂X onto $\partial X'$. Hence f can be extended to a homeomorphism of X onto X' . These extensions on different 1-tiles paste together to a continuous map $f: S^2 \rightarrow S^2$ that is cellular and is compatible with the given labeling. Moreover, $f|_X$ is orientation-preserving for each 1-tile X as follows from the fact that the labeling is orientation-preserving. By Lemma 5.2 (i) and (ii) the map f is a postcritically-finite branched covering map. This shows that a map with the stated properties exists.

To show uniqueness suppose that $g: S^2 \rightarrow S^2$ is another such map. Then for each cell $\tau \in \mathcal{D}^1$, the maps $f|_\tau$ and $g|_\tau$ are homeomorphisms of τ onto $L(\tau) \in \mathcal{D}^0$. Hence $\varphi_\tau = (g|_\tau)^{-1} \circ (f|_\tau)$ is a homeomorphism of τ onto itself. The family φ_τ , $\tau \in \mathcal{D}^1$, of these homeomorphisms is obviously compatible under inclusions: if $\sigma, \tau \in \mathcal{D}^1$ and $\sigma \subset \tau$, then $\varphi_\tau(p) = \varphi_\sigma(p)$ for all $p \in \sigma$.

Using this we can define a map $\varphi: S^2 \rightarrow S^2$ as follows. For $p \in S^2$ pick $\tau \in \mathcal{D}^1$ with $p \in \tau$. Then set $\varphi(p) = \varphi_\tau(p)$. The compatibility properties of the homeomorphisms φ_τ imply that φ is well-defined. Indeed, suppose that τ, τ' are cells in \mathcal{D}^1 with $p \in \tau \cap \tau'$. There exists a unique cell $\sigma \in \mathcal{D}^1$ with $p \in \text{int}(\sigma)$. It follows from Lemma 4.3 (ii) that $\sigma \subset \tau \cap \tau'$. Hence

$$\varphi_\tau(p) = \varphi_\sigma(p) = \varphi_{\tau'}(p).$$

It is clear that $g \circ \varphi = f$. Moreover, $\varphi|_\tau = \varphi_\tau$ is a homeomorphism of τ onto itself whenever $\tau \in \mathcal{D}^1$. This implies that φ is continuous as there only finitely many cells in \mathcal{D}^1 , and that φ is surjective. The map φ is also injective as follows from the facts that $\varphi(\text{int}(\tau)) = \text{int}(\tau)$ and that $\varphi|_\tau$ is injective for each $\tau \in \mathcal{D}^1$, and that S^2 is the disjoint union of the sets $\text{int}(\tau)$, $\tau \in \mathcal{D}^1$.

Hence $\varphi: S^2 \rightarrow S^2$ is continuous and bijective, and so a homeomorphism of S^2 onto itself. Note that f and g agree on the set \mathbf{V}^1 of 1-vertices. Hence φ is the identity on \mathbf{V}^1 . Moreover, the sets of postcritical points of f and g are contained in the set of 0-vertices and hence in \mathbf{V}^1 . So Thurston equivalence of f and g will follow, if we can show that φ is isotopic to id_{S^2} rel. \mathbf{V}^1 .

This again follows from a procedure based on successive extension on skeleta of \mathcal{D}^1 . Indeed, let \mathbf{E}^1 be the set of 1-edges, $E^1 = \bigcup \{e : e \in \mathbf{E}^1\}$ be the 1-skeleton of \mathcal{D}^1 , and $e \in \mathbf{E}^1$ be an arbitrary 1-edge. Since φ is the identity on vertices in \mathcal{D}^1 , the map $\varphi|_e = \varphi_e$ is isotopic to id_e rel. ∂e . Then these isotopies on edges paste together to an isotopy of $\varphi|_{E^1}$ to id_{E^1} rel. \mathbf{V}^1 . If X is a tile in \mathcal{D}^1 , then this isotopy is defined on $\partial X \subset E^1$, and we can extend it to an isotopy of a homeomorphism

on X that agrees with $\varphi|_{\partial X}$ on ∂X to id_X . These extensions on tiles X paste together to an isotopy $\Phi: S^2 \times [0, 1] \rightarrow S^2$ rel. \mathbf{V}^1 such that $\Phi(\cdot, 1) = \text{id}_{S^2}$ and $\tilde{\varphi}|_{E^1} = \varphi|_{E^1}$, where $\tilde{\varphi} := \Phi(\cdot, 0)$.

For each tile $X \in \mathcal{D}^1$ the maps $\varphi|_X$ and $\tilde{\varphi}|_X$ are homeomorphisms of X onto itself that agree on $\partial X \subset E^1$. As we have seen in the discussion before the statement of the proposition, this implies that $\varphi|_X$ and $\tilde{\varphi}|_X$ are isotopic rel. ∂X . By pasting these isotopies on tiles together, we can find an isotopy $\Psi: S^2 \times [0, 1] \rightarrow S^2$ rel. E^1 with $\Psi(\cdot, 0) = \varphi$ and $\Psi(\cdot, 1) = \tilde{\varphi}$. The concatenation of the isotopies Ψ and Φ gives the desired isotopy rel. \mathbf{V}^1 between φ and id_{S^2} . \square

As we know from Section 6, every Thurston map $f: S^2 \rightarrow S^2$ arises from cell decompositions \mathcal{D}^1 and \mathcal{D}^0 of S^2 as in the last proposition. This gives a useful description of a Thurston map in combinatorial terms. If one wants to study the dynamics of f , one is interested in the cell decompositions \mathcal{D}^n obtained from pulling back \mathcal{D}^0 by f^n as in Lemma 5.4. In general, in order to determine the combinatorics of the whole sequence \mathcal{D}^n , $n \in \mathbb{N}_0$ (i.e., the inclusion and intersection patterns of cells on *all* levels), is not enough to just know the pair $(\mathcal{D}^0, \mathcal{D}^1)$ and the labeling $\tau \in \mathcal{D}^1 \mapsto f(\tau) \in \mathcal{D}^0$, but one also needs specific information on the *pointwise* mapping behavior of f on the cells in \mathcal{D}^1 . Indeed, suppose g is another map that is cellular for $(\mathcal{D}^0, \mathcal{D}^1)$ and induces the same labeling as f , i.e., $f(\tau) = g(\tau)$ for all $\tau \in \mathcal{D}^1$. Let $\tilde{\mathcal{D}}^n$ be the cell decomposition of S^2 obtained from \mathcal{D}^0 by pulling back by g^n . Then one can show (by an argument very similar to the considerations in the proof of Lemma 12.9 below) that \mathcal{D}^n and $\tilde{\mathcal{D}}^n$ are isomorphic cell complexes for *fixed* $n \in \mathbb{N}_0$ (see Definition 12.2). In contrast, the intersection patterns of corresponding cells in \mathcal{D}^n and $\tilde{\mathcal{D}}^n$ on *distinct* levels n may be quite different.

The situation changes if $(\mathcal{D}^1, \mathcal{D}^0)$ is a cellular Markov partition for f , because then the combinatorics of \mathcal{D}^n is completely determined by $(\mathcal{D}^1, \mathcal{D}^0)$ and the combinatorial data given by the labeling $\tau \in \mathcal{D}^1 \mapsto f(\tau) \in \mathcal{D}^0$ (see Remark 4.11). This suggests that if one wants to study Thurston maps as given by Proposition 12.3 from a purely combinatorial point of view, then one should add the additional assumption that \mathcal{D}^1 is a refinement of \mathcal{D}^0 . We will restrict ourselves to the case where \mathcal{D}^0 contains only two tiles. Then all the relevant assumptions can be condensed into the following definition.

Definition 12.4 (Two-tile subdivision rules). Let S^2 be a 2-sphere. A *two-tile subdivision rule* for S^2 is a triple $(\mathcal{D}^1, \mathcal{D}^0, L)$ of cell decompositions \mathcal{D}^0 and \mathcal{D}^1 of S^2 and an orientation-preserving labeling

$L: \mathcal{D}^0 \rightarrow \mathcal{D}^1$. We assume that the cell decompositions satisfy the following conditions:

- (i) \mathcal{D}^0 contains precisely two 0-tiles.
- (ii) \mathcal{D}^1 is a refinement of \mathcal{D}^0 , and \mathcal{D}^1 contains more than two tiles.
- (iii) If k is the number of 0-vertices, then $k \geq 3$ and every tile in \mathcal{D}^1 is a k -gon.
- (iv) The length of the cycle of every 1-vertex is even.

Our concept of a two-tile subdivision rule is a special case of the more general concept of a *subdivision rule*. See [BS, CFP01, CFP06a, CFKP, Me02] for work related to subdivision rules. The reason for the name *two-tile* subdivision rule is that the data given by $(\mathcal{D}^1, \mathcal{D}^0)$ determines how the *two* 0-tiles are subdivided by the cells in \mathcal{D}^1 , and this together with the labeling L can be used to create a sequence of cell decomposition \mathcal{D}^n where each cell $\tau \in \mathcal{D}^1$ is subdivided by the cells in \mathcal{D}^2 in the same as as the cell $L(\tau) \in \mathcal{D}^0$ is subdivided by the 1-cells, etc. Our definition is tailored to generate Thurston maps, so a more accurate term would have been a “two-tile subdivision rule generating a Thurston map”, but we chose the shorter term for brevity.

Conversely, suppose $f: S^2 \rightarrow S^2$ is a Thurston map with $k := \text{post}(f) \geq 3$, and $\mathcal{C} \subset S^2$ is an f -invariant curve with $\text{post}(f) \subset \mathcal{C}$. If $\mathcal{D}^0 = \mathcal{D}^0(f, \mathcal{C})$, $\mathcal{D}^1 = \mathcal{D}^1(f, \mathcal{C})$, and $L: \mathcal{D}^1 \rightarrow \mathcal{D}^0$ is the (orientation-preserving) labeling induced by f , then $(\mathcal{D}^1, \mathcal{D}^0, L)$ is a two-tile subdivision rule. This immediately follows from Proposition 6.1 and Proposition 11.1.

Let \mathcal{D}^0 be a cell decomposition of S^2 with precisely two tiles X and Y . Then necessarily $\partial X = \partial Y$. The set $\mathcal{C} := \partial X = \partial Y$ is a Jordan curve which we call *the Jordan curve of \mathcal{D}^0* . Then \mathcal{C} is the 1-skeleton of \mathcal{D}^0 and all vertices and edges of \mathcal{D}^0 lie on \mathcal{C} . If k is the number of these vertices on \mathcal{C} and \mathcal{D}^1 is another cell decomposition of S^2 , then a Thurston map that is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$ can only exist if each tile in \mathcal{D}^1 is a k -gon, i.e., it contains exactly k vertices and edges in its boundary. Moreover, the length of each vertex cycle in \mathcal{D}^1 has to be even, because it must be an integer multiple of the length of a vertex cycle in \mathcal{D}^0 which is always equal to 2. This motivated conditions (iii) and (iv) in Definition 12.4.

The next proposition immediately follows from Proposition 12.3 and gives a large supply of Thurston maps.

Proposition 12.5. *Let $(\mathcal{D}^1, \mathcal{D}^0, L)$ be a two-tile subdivision rule on S^2 . Then there exists a Thurston map $f: S^2 \rightarrow S^2$ that is cellular for*

$(\mathcal{D}^0, \mathcal{D}^1)$ and is compatible with the labeling L . The map f is unique up to Thurston equivalence.

Moreover, the Jordan curve \mathcal{C} of \mathcal{D}^0 is f -invariant and contains the set $\text{post}(f)$.

Proof. The first part is just a special case of Proposition 12.3. Note that f is a Thurston map; indeed, the number of 1-tiles is equal to $2 \deg(f)$, and also > 2 by condition (ii) in Definition 12.4 (ii). So $\deg(f) \geq 2$.

Since \mathcal{D}^1 is a refinement of \mathcal{D}^0 , the 1-skeleton \mathcal{C} of \mathcal{D}^0 is contained in the 1-skeleton of \mathcal{D}^1 . Moreover, since f is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$, this map sends the 1-skeleton of \mathcal{D}^1 into the 1-skeleton of \mathcal{D}^0 . Hence $f(\mathcal{C}) \subset \mathcal{C}$, and so \mathcal{C} is f -invariant. Each postcritical point of f is a vertex of \mathcal{D}^0 and hence contained in \mathcal{C} . \square

If the map f is as in Proposition 12.5, then we say that it *realizes* the two-tile subdivision rule. If, as usual, \mathbf{V}^0 denotes the set of vertices of \mathcal{D}^0 and \mathbf{V}^1 the set of vertices of \mathcal{D}^1 , we then have $\text{crit}(f) \subset \mathbf{V}^1$ and $\text{post}(f) \subset \mathbf{V}^0$ (see Lemma 5.2). Since the length of each cycle in \mathcal{D}^0 is 2, a vertex v in \mathbf{V}^1 is a critical point of f if and only if the length of the cycle of v in \mathcal{D}^1 is ≥ 4 (see Remark 5.3). Hence if f and g both realize the subdivision rule, then $\text{crit}(f) = \text{crit}(g) \subset \mathbf{V}^1$. Moreover, since the orbit of any point in \mathbf{V}^1 is completely determined by the labeling, we then also have $\text{post}(f) = \text{post}(g) \subset \mathbf{V}^0$.

Theorem 1.2 implies that every expanding Thurston map f with $\#\text{post}(f) \geq 3$ has an iterate $F = f^n$ that is obtained from a two-tile subdivision rule as in the previous proposition.

If one wants to discuss specific examples of Thurston maps that realize a given two-tile subdivision rule $(\mathcal{D}^1, \mathcal{D}^0, L)$, then it is convenient to represent the relevant data in a compressed form. As we will see, the information on L is completely determined by a pair of corresponding positively-oriented flags in \mathcal{D}^1 and \mathcal{D}^0 . See Lemma 12.7 below for a precise statement. To describe labelings for pairs $(\mathcal{D}^1, \mathcal{D}^0)$ as in Definition 12.4 we proceed in the manner discussed after Definition 12.1 and choose a more general index set \mathcal{L} for the labeling of the elements in \mathcal{D}^0 and \mathcal{D}^1 .

To set this up, it is useful to introduce some terminology first. Let X be a closed Jordan region on S^2 with distinct points $v_0, \dots, v_{k-1}, v_k = v_0$, $k \geq 3$, on its boundary. Here we use the cyclic group $\mathbb{Z}_k = \{0, 1, \dots, k-1\} = \mathbb{Z}/k\mathbb{Z}$ as an index set.

Suppose that the points v_0, \dots, v_{k-1} are indexed such that if we start at v_0 and run through ∂X with suitable orientation, then the points v_0, \dots, v_{k-1} are traversed in successive order. If this is true and if with

this orientation of ∂X the region X lies on the left, then we call the points v_1, \dots, v_k in *cyclic* order on ∂X , and otherwise, if X lies on the right, in *anti-cyclic* order on ∂X . If the points v_0, \dots, v_{k-1} are in cyclic or anti-cyclic order on ∂X , then ∂X is decomposed into unique arcs e_0, \dots, e_{k-1} ; here e_l for $l \in \mathbb{Z}_k$ is the unique subarc of ∂X that has the endpoints v_l and v_{l+1} , but does not contain any other of the points v_i , $i \in \mathbb{Z}_k \setminus \{l, l+1\}$. We say that the arcs e_0, \dots, e_{k-1} are in *cyclic* or *anti-cyclic* order on ∂X , if this is true for the points v_0, \dots, v_{k-1} , respectively.

Let \mathcal{D} be a cell decomposition of S^2 . A chain of tiles X_1, \dots, X_N in \mathcal{D} is called an *e-chain* if for $i = 1, \dots, N-1$ we have $X_i \neq X_{i+1}$ and there exists an edge e_i in \mathcal{D} with $e_i \subset \partial X_i \cap \partial X_{i+1}$. The *e-chain joins* the tiles X and Y if $X_1 = X$ and $X_N = Y$. If X is an arbitrary tile in \mathcal{D} , then every tile Y in \mathcal{D} can be joined to X by an *e-chain*. This follows from the fact that the union of the tiles Y that can be joined to X is equal to S^2 ; indeed, this union is a nonempty closed set, and it is also open, as follows from Lemma 5.1 (iv) and (v). Hence the union is all of S^2 .

Similarly as in Lemma 6.2, we will label the tiles in \mathcal{D} by the two symbols **b** and **w**, representing the two colors “black” and “white”, respectively. So then each tile in \mathcal{D} will carry one of these colors.

The following lemma will be the basis for the construction of labelings.

Lemma 12.6. *Let \mathcal{D} be a cell decomposition of S^2 , and denote by \mathbf{V} the set of vertices, by \mathbf{E} the set of edges, and by \mathbf{X} the set of tiles in \mathcal{D} . Suppose that the length of the cycle of every vertex in \mathcal{D} is even and that there exists $k \geq 3$ such that every tile in \mathbf{X} is a k -gon.*

Then for each positively-oriented flag (c_0, c_1, c_2) in \mathcal{D} there exist maps $L_{\mathbf{V}}: \mathbf{V} \rightarrow \mathbb{Z}_k$, $L_{\mathbf{E}}: \mathbf{E} \rightarrow \mathbb{Z}_k$, and $L_{\mathbf{X}}: \mathbf{X} \rightarrow \{\mathbf{b}, \mathbf{w}\}$ with the following properties:

- (i) $L_{\mathbf{V}}(p_0) = 0$, where $c_0 = \{p_0\}$, $L_{\mathbf{E}}(c_1) = 0$, and $L_{\mathbf{X}}(c_2) = \mathbf{w}$,
- (ii) if $X, Y \in \mathbf{X}$ are two distinct tiles with a common edge on their boundaries, then $L_{\mathbf{X}}(X) \neq L_{\mathbf{X}}(Y)$,
- (iii) if X is an arbitrary tile in \mathbf{X} , and v_0, \dots, v_{k-1} are the vertices on its boundary indexed by \mathbb{Z}_k , then we can choose the indices of the vertices so that $L_{\mathbf{V}}(v_i) = i$ for each $i \in \mathbb{Z}_k$, and such that the order of the vertices on ∂X is cyclic if $L_{\mathbf{X}}(X) = \mathbf{w}$ and anti-cyclic if $L_{\mathbf{X}}(X) = \mathbf{b}$,
- (iv) if $e \in \mathbf{E}$ and $l = L_{\mathbf{E}}(e)$, then $L_{\mathbf{V}}(\partial e) = \{l, l+1\}$,

- (v) if X is an arbitrary tile in \mathbf{X} , and e_0, \dots, e_{k-1} are the edges on its boundary indexed by \mathbb{Z}_k , then we can choose the indices of the edges so that $L_{\mathbf{E}}(e_i) = i$ for each $i \in \mathbb{Z}_k$, and such that the order of the edges on ∂X is cyclic if $L_{\mathbf{X}}(X) = \mathbf{w}$ and anti-cyclic if $L_{\mathbf{X}}(X) = \mathbf{b}$,
- (vi) if (τ_0, τ_1, τ_2) is a flag in \mathcal{D} , then the flag is positively-oriented if and only if there exists $l \in \mathbb{Z}_k$ such that $L_{\mathbf{V}}(\tau_0) = \{l\}$, $L_{\mathbf{E}}(\tau_1) = l$, $L_{\mathbf{X}}(\tau_2) = \mathbf{w}$, or $L_{\mathbf{V}}(\tau_0) = \{l\}$, $L_{\mathbf{E}}(\tau_1) = l - 1$, $L_{\mathbf{X}}(\tau_2) = \mathbf{b}$.

The maps $L_{\mathbf{V}}$, $L_{\mathbf{E}}$, and $L_{\mathbf{X}}$ are uniquely determined by the properties (i)–(iv).

Condition (ii) says that one of the tiles containing an edge is “black” and the other is “white”. Moreover, by (iii) and (v) we can index the vertices v and edges e on the boundary of a tile by the label $L_{\mathbf{X}}(v) \in \mathbb{Z}_k$ and $L_{\mathbf{E}}(e) \in \mathbb{Z}_k$, respectively, so that the vertices and edges of a white tile are in cyclic order and the ones on a black tile are in anti-cyclic order. By (iv) the label $L_{\mathbf{E}}(e)$ of an edge $e \in \mathbf{E}$ is determined by the labels $L_{\mathbf{V}}(u)$ and $L_{\mathbf{V}}(v)$ of the two endpoints of e (here it is important that $k \geq 3$).

Proof. We first establish the following statement.

Claim. Suppose that $J \subset S^2$ is a Jordan curve that does not contain any vertex (in \mathcal{D}) and has the property that for every edge e the intersection $e \cap J$ is either empty, or e meets both components of $S^2 \setminus J$ and $e \cap J$ consists of a single point. Then J meets an even number of edges.

To see this pick one of the complementary components U of $S^2 \setminus J$, and let d_1, \dots, d_n , $n \in \mathbb{N}_0$, be the length of the cycles of the vertices contained in U (for $n = 0$ we consider this as an empty list). Let \mathbf{E}_J be the set of all edges that meet J and \mathbf{E}_U be the set of all edges contained in U . From our assumption on the intersection property of J with edges it follows that an edge is contained in U if and only if its two endpoints are in U , and it meets J if and only if one endpoint is in U and the other in $S^2 \setminus \overline{U}$. Hence

$$d_1 + \dots + d_n = \#\mathbf{E}_J + 2\#\mathbf{E}_U,$$

because the sum on the left hand side counts every edge in \mathbf{E}_J once, and every edge in \mathbf{E}_U twice. Since all the numbers d_1, \dots, d_n are even, it follows that the number $\#\mathbf{E}_J$ of edges that J meets is also even, proving the claim.

To show existence and uniqueness of the map $L_{\mathbf{X}}$ we proceed as follows. For every tile Y there exists an e -chain $Y_0 = c_2, \dots, Y_N = Y$ of

tiles joining the “base tile” c_2 to Y . We put $L_{\mathbf{X}}(Y) = \mathbf{w}$ or $L_{\mathbf{X}}(Y) = \mathbf{b}$ depending on whether N is even or odd. It is clear that if this is well-defined, then it is the unique choice for $L_{\mathbf{X}}(Y)$. This follows from the normalization (i) and that fact that by (ii) the labels of tiles have to alternate along an e -chain.

To see that $L_{\mathbf{X}}$ is well-defined it is enough to show that if an e -chain X_0, X_1, \dots, X_N forms a cycle, i.e., if $X_0 = X_N$, then N is even. To prove this we may make the additional assumption that $N \geq 3$ and that the chain is *simple*, i.e., that the tiles X_1, \dots, X_N are all distinct.

We can choose edges e_i for $i = 1, \dots, N$ such that $e_i \subset \partial X_{i-1} \cap \partial X_i$. Then the edges e_1, \dots, e_N are all distinct. For otherwise, $e_i = e_j$ for some $1 \leq i < j \leq N$. Then $e_i = e_j$ is contained in the boundary of the tiles $X_{i-1}, X_i, X_{j-1}, X_j$ which is impossible, because three of these tiles must be distinct (note that $N \geq 3$).

We now construct a Jordan curve J as follows. For each edge e_i pick a point $x_i \in \text{int}(e_i)$. Moreover, for $i = 1, \dots, N$, we can choose an arc $\alpha_i \subset X_i$ with endpoints x_i and x_{i+1} such that $\text{int}(\alpha_i) \subset \text{int}(X_i)$. Here $x_{N+1} := x_1$. Then $J = \alpha_1 \cup \dots \cup \alpha_N$ is a Jordan curve that has properties as in the claim above. The curve J meets the edges e_1, \dots, e_N and no others. Hence N is even. Thus $L_{\mathbf{X}}$ is well-defined, and it satisfies property (ii) and is normalized as in (i).

To show the existence of $L_{\mathbf{V}}$ it is useful to quickly recall some basic definitions from the homology and cohomology of chain complexes. Denote by \mathbf{E}_o the set of oriented edges in \mathcal{D} . Let $C(\mathbf{X})$ and $C(\mathbf{E}_o)$ be the free modules over \mathbb{Z}_k generated by the sets \mathbf{X} and \mathbf{E}_o , respectively. So $C(\mathbf{E}_o)$, for example, is just the set of formal finite sums $\sum a_i e_i$, where $a_i \in \mathbb{Z}_k$ and $e_i \in \mathbf{E}_o$. Note that in contrast to other commonly used definitions of chain complexes we have $e + \tilde{e} \neq 0$ if e and \tilde{e} are the same edges with opposite orientation.

There is a unique boundary operator $b: C(\mathbf{X}) \rightarrow C(\mathbf{E}_o)$ that is a module homomorphism and satisfies

$$bX := b(X) = \sum_{e \in \partial X} e$$

for each tile X , where the sum is extended over all oriented edges $e \in \partial X$ so that X lies on the left of e .

Let e be an oriented edge and X be the unique tile with $e \subset \partial X$ that is on the left of e . We put $\alpha(e) = 1 \in \mathbb{Z}_k$ or $\alpha(e) = -1 \in \mathbb{Z}_k$ depending on whether $L_{\mathbf{X}}(X) = \mathbf{w}$ (X is a white tile) or $L_{\mathbf{X}}(X) = \mathbf{b}$ (X is a black tile). If e and \tilde{e} are the same edges with opposite orientation, then $\alpha(e) + \alpha(\tilde{e}) = 0$ as follows from property (ii) of $L_{\mathbf{X}}$.

The map α can be uniquely extended to a homomorphism $\alpha: C(\mathbf{E}_o) \rightarrow \mathbb{Z}_k$. In the language of cohomology it is a “cochain”. This cochain α is a cocycle, i.e.,

$$(12.1) \quad \alpha(bX) = \sum_{e \subset \partial X} \alpha(e) = \pm k = 0 \in \mathbb{Z}_k$$

for every tile X , considered as one of the generators of $C(\mathbf{X})$. Indeed, by our convention on the orientation of edges $e \subset \partial X$ in the above sum, for each such edge we get the same contribution $\alpha(e)$ in the sum, and so, since X has k edges, the sum is equal to $\pm k = 0 \in \mathbb{Z}_k$.

Consider an arbitrary closed edge path consisting of the oriented edges e_1, \dots, e_n ; so the terminal point of e_i is the initial point of e_{i+1} for $i = 1, \dots, n$, where $e_{n+1} := e_1$. We claim that

$$(12.2) \quad \sum_{i=1}^n \alpha(e_i) = 0.$$

Essentially, this is a consequence of the fact that we have $H^1(S^2, \mathbb{Z}_k) = 0$ for the first cohomology group of S^2 with coefficients in \mathbb{Z}_k . This implies that the cocycle α is a coboundary.

We will present a simple direct argument. To show (12.2) it is clearly enough to establish this for simple closed edge paths, i.e., for closed edge paths where the union of the edges forms a Jordan curve $J \subset S^2$. Let U be the complementary component of $S^2 \setminus J$ so that U lies on the left if we traverse J according to the orientation given by the edges e_i . If X_1, \dots, X_M are all the tiles contained in \overline{U} , then

$$b(X_1 + \dots + X_M) = \sum_{e \subset \overline{U}} e,$$

where the sum is extended over oriented edges contained in \overline{U} . Each edge on J is equal to one of the edges e_i and it appears in the above sum exactly once and with the same orientation as e_i . All other edges in \overline{U} appear twice and with opposite orientations. Hence by (12.1),

$$\sum_{i=1}^n \alpha(e_i) = \sum_{e \subset \overline{U}} \alpha(e) = \sum_{i=1}^M \alpha(bX_i) = 0.$$

We now define $L_{\mathbf{V}}: \mathbf{V} \rightarrow \mathbb{Z}_k$ as follows. For $v \in \mathbf{V}$ we can pick an edge path consisting of the oriented edges e_1, \dots, e_n that joins the base point p_0 to v (this list of edges may be empty if $v = p_0$). The existence of such an edge path follows from the connectedness of the 1-skeleton

of \mathcal{D} (see Lemma 5.1 (vi)). Put

$$(12.3) \quad L_{\mathbf{V}}(v) = \sum_{i=1}^n \alpha(e_i).$$

This is well-defined, because we have (12.2) for every closed edge path; we also have the normalization $L_{\mathbf{V}}(p_0) = 0$.

The definition of $L_{\mathbf{V}}$ implies that if e is an oriented edge, and u is the initial and v the terminal point of e , then

$$(12.4) \quad L_{\mathbf{V}}(v) = L_{\mathbf{V}}(u) + \alpha(e).$$

This means that if we go from the initial point u of e to the terminal point v , then the value of $L_{\mathbf{V}}$ is increased by 1 or decreased by -1 depending on whether the tile on the left of e is white or black. The desired property (iii) of $L_{\mathbf{V}}$ immediately follows from this.

This shows existence of $L_{\mathbf{V}}$. Conversely, every function $L_{\mathbf{V}}$ with property (iii) must satisfy (12.4). Together with the normalization $L_{\mathbf{V}}(p_0) = 0$ this implies that $L_{\mathbf{V}}$ is given by the formula (12.3), and so we have uniqueness.

To define $L_{\mathbf{E}}$ note that if $e \in \mathbf{E}$, then by (ii) we can choose a unique orientation for e such that the tile on the left is “white”, and the one on the right is “black”. If u is the initial and v the terminal point of e according to this orientation, and $L_{\mathbf{V}}(u) = l \in \mathbb{Z}_k$, then $L_{\mathbf{V}}(v) = l + 1$. Now set $L_{\mathbf{E}}(e) := l$. Then $L_{\mathbf{E}}$ has property (iv). Moreover, we also have the normalization (i) for $L_{\mathbf{E}}$; indeed, if c_1 is oriented so that p_0 is the initial point of c_1 , then c_2 lies on the left of c_1 , because the flag (c_0, c_1, c_2) is positively-oriented. Since $L_{\mathbf{X}}(c_2) = \mathbf{w}$, the tile c_2 is white and so $L_{\mathbf{E}}(e) = L_{\mathbf{V}}(p_0) = 0$. Uniqueness of $L_{\mathbf{E}}$ follows from (iii) and the uniqueness of $L_{\mathbf{V}}$.

We have proved (i)–(iv) and the uniqueness statement. It remains to establish (v) and (vi).

To show (v) let $X \in \mathbf{X}$ be arbitrary. Then by (iii) we can assume that the indexing of the k vertices v_0, \dots, v_{k-1} on ∂X is such that $L_{\mathbf{V}}(v_i) = i$ for all $i \in \mathbb{Z}_k$, and that v_0, \dots, v_{k-1} are met in successive order if we traverse ∂X . This implies that for each $i \in \mathbb{Z}_k$ there exists a unique edge $e_i \subset \partial X$ in \mathcal{D} with endpoints v_i and v_{i+1} . Hence by (iv) we have $L_{\mathbf{E}}(e_i) = i$. Moreover, by (iii) the edges e_0, \dots, e_{k-1} are in cyclic or anti-cyclic order on ∂X depending on whether $L_{\mathbf{X}}(X) = \mathbf{w}$ or $L_{\mathbf{X}}(X) = \mathbf{b}$. So (v) holds.

Finally, to see that (vi) is true, let (τ_0, τ_1, τ_2) be a flag in \mathcal{D} . Then $\tau_0 = \{u\}$ for some $u \in \mathbf{V}$. The vertex u is the initial point of the oriented edge τ_1 . Let $v \in \mathbf{V}$ be the terminal point of τ_1 , and define $l = L_{\mathbf{V}}(u)$.

Depending on whether the flag is positively- or negatively-oriented, the vertex v follows u in cyclic or anti-cyclic order on $\partial\tau_2$. So if the flag is positively-oriented, then by property (iii) we have $L_{\mathbf{V}}(v) = l + 1$ if $L_{\mathbf{X}}(\tau_2) = \mathbf{w}$ and $L_{\mathbf{V}}(v) = l - 1$ if $L_{\mathbf{X}}(\tau_2) = \mathbf{b}$. Property (iv) implies that $L_{\mathbf{E}}(\tau_1) = l$ if $L_{\mathbf{X}}(\tau_2) = \mathbf{w}$ and $L_{\mathbf{E}}(\tau_1) = l - 1$ if $L_{\mathbf{X}}(\tau_2) = \mathbf{b}$.

So if (τ_0, τ_1, τ_2) is positively-oriented, then the cells in this flag carry the labels l, l, \mathbf{w} , or $l, l - 1, \mathbf{b}$, respectively.

Similarly, if (τ_0, τ_1, τ_2) is negatively-oriented, then we get the labels $l, l - 1, \mathbf{w}$, or l, l, \mathbf{b} for the cells in the flag. Statement (vi) follows from this. \square

Lemma 12.7. *Let $(\mathcal{D}^1, \mathcal{D}^0)$ be a pair of cell decompositions of S^2 satisfying conditions (i)–(iv) in Definition 12.4, and let (c'_0, c'_1, c'_2) and (c_0, c_1, c_2) be positively-oriented flags in \mathcal{D}^1 and \mathcal{D}^0 , respectively. Then there exists a unique orientation-preserving labeling $L: \mathcal{D}^1 \rightarrow \mathcal{D}^0$ with $(L(c'_0), L(c'_1), L(c'_2)) = (c_0, c_1, c_2)$.*

In particular, $(\mathcal{D}^1, \mathcal{D}^0, L)$ is a two-tile subdivision rule.

Proof. For $i = 0, 1$ denote by $\mathbf{V}^i, \mathbf{E}^i, \mathbf{X}^i$ the set of vertices, edges, and tiles of \mathcal{D}^i , respectively.

To describe the labeling for $(\mathcal{D}^1, \mathcal{D}^0)$ we proceed in the manner discussed after Definition 12.1 and choose a particular index set \mathcal{L} for the labeling of the elements in \mathcal{D}^0 and \mathcal{D}^1 .

We let \mathcal{L} be the set that consists of two disjoint copies of \mathbb{Z}_k (one will be for the vertices, and one for the edges), and the set $\{\mathbf{b}, \mathbf{w}\}$, where again we think of \mathbf{w} representing “white” and \mathbf{b} representing “black”.

We assign to $c_2 \in \mathbf{X}^0$ the color “white”, and “black” to the other tile in \mathbf{X}^0 . We assign $0 \in \mathbb{Z}_k$ to the 0-vertex $v_0 \in c_0$. Then there is a unique way to assign labels in \mathbb{Z}_k to the other vertices on $\mathcal{C} := \partial c_2$ (and the corresponding cells of dimension 0) such that if v_0, v_1, \dots, v_{k-1} are the vertices indexed by their label, then they are in cyclic order on \mathcal{C} as considered as the boundary of the white 0-tile and in anti-cyclic order for the black 0-tile. Each 0-edge e is an arc on \mathcal{C} with endpoints v_l and v_{l+1} for a unique $l \in \mathbb{Z}_k$. We label e by l (where l is thought of to belong to the second copy of \mathbb{Z}_k). Since (c_0, c_1, c_2) is a positively-oriented flag, and v_0 is the initial point of c_1 , the edge c_1 has the label 0. All this is just a special case of Lemma 12.6. If in this way we assign to each element in \mathcal{D}^0 a label in \mathcal{L} , we get a bijection $\psi: \mathcal{D}^0 \rightarrow \mathcal{L}$. Note that if (τ_0, τ_1, τ_2) is any positively-oriented flag in \mathcal{D}^0 , then its image under ψ has the form (l, l, \mathbf{w}) or $(l, l - 1, \mathbf{b})$ for some $l \in \mathbb{Z}_k$ (cf. Lemma 12.6 (v)).

For \mathcal{D}^1 we invoke Lemma 12.6 directly to set up a suitable map $\varphi: \mathcal{D}^1 \rightarrow \mathcal{L}$. Since \mathcal{D}^1 satisfies the conditions of Lemma 12.6, we can

find maps $L_{\mathbf{V}}: \mathbf{V}^1 \rightarrow \mathbb{Z}_k$, $L_{\mathbf{E}}: \mathbf{E}^1 \rightarrow \mathbb{Z}_k$, and $L_{\mathbf{X}}: \mathbf{X}^1 \rightarrow \{\mathbf{b}, \mathbf{w}\}$ with the properties (ii)–(iv) stated in the lemma and the normalizations $L_{\mathbf{V}}(v'_0) = 0$, where $c'_0 = \{v'_0\}$, $L_{\mathbf{E}}(c'_1) = 0$, and $L_{\mathbf{X}}(c'_2) = \mathbf{w}$. The maps $L_{\mathbf{V}}$, $L_{\mathbf{E}}$, $L_{\mathbf{X}}$ induce a unique map $\varphi: \mathcal{D}^1 \rightarrow \mathcal{L}$ such that $\varphi(c) = L_{\mathbf{X}}(c)$ if c is a 1-tile, $\varphi(c) = L_{\mathbf{E}}(c)$ if c is a 1-edge, and $\varphi(c) = L_{\mathbf{V}}(v)$ if $c = \{v\}$ consists of a 1-vertex v .

Now define $L := \psi^{-1} \circ \varphi: \mathcal{D}^1 \rightarrow \mathcal{D}^0$. The map L assigns to each 1-cell c the unique 0-cell that has the same dimension as c and carries the same label in \mathcal{L} as c .

It follows immediately from the properties of the maps ψ and φ that L preserves dimensions, respects inclusions, and is injective on cells. Hence L is a labeling according to Definition 12.1. By our normalizations the map L sends the flag (c'_0, c'_1, c'_2) to (c_0, c_1, c_2) .

Moreover, L is orientation-preserving. Indeed, φ maps the cells τ_0, τ_1, τ_2 in a positively-oriented flag in \mathcal{D}^1 to l, l, \mathbf{w} , or to $l, l-1, \mathbf{b}$, respectively, where $l \in \mathbb{Z}_k$. These triples correspond to positively-oriented flags in \mathcal{D}^0 . It follows that L has the desired properties.

To show uniqueness, we reverse the process. Given L with the stated properties, we use the same map $\psi: \mathcal{D}^0 \rightarrow \mathcal{L}$ as above and define maps $L_{\mathbf{V}}: \mathbf{V}^1 \rightarrow \mathbb{Z}_k$, $L_{\mathbf{E}}: \mathbf{E}^1 \rightarrow \mathbb{Z}_k$, $L_{\mathbf{X}}: \mathbf{X}^1 \rightarrow \{\mathbf{b}, \mathbf{w}\}$ such that $L_{\mathbf{X}}(c) = (\psi \circ L)(c)$ if c is a 1-tile, $L_{\mathbf{E}}(c) = (\psi \circ L)(c)$ if c is a 1-edge, and such that $L_{\mathbf{V}}(v) = (\psi \circ L)(c)$ if $c = \{v\}$ consists of a 1-vertex v .

Then we have normalizations $L_{\mathbf{V}}(v'_0) = 0$, $L_{\mathbf{E}}(c'_1) = 0$, and $L_{\mathbf{X}}(c'_2) = \mathbf{w}$ as in Lemma 12.6 (i). If we can show that $L_{\mathbf{V}}$, $L_{\mathbf{E}}$, $L_{\mathbf{X}}$ have the properties (ii)–(iv) in Lemma 12.6, then the uniqueness of L will follow from the corresponding uniqueness statement in this lemma.

To see this let $e \in \mathcal{D}^1$ be arbitrary, and $X, Y \in \mathcal{D}^1$ be the two tiles that contain e in its boundary. Let $u, v \in \mathbf{V}^1$ be the two endpoints of e . We may assume that notation is chosen so that the flag $(\{u\}, e, X)$ is positively-oriented. Then $(\{v\}, e, Y)$ is also positively-oriented. It follows that the images of these flags under L are positively-oriented. Since L is injective on cells, and so $L(u) \neq L(v)$, this implies that $L(X) \neq L(Y)$. So $L(X)$ and $L(Y)$ carry different colors (given by ψ) which implies that X and Y also carry different colors by definition of $L_{\mathbf{X}}$. Hence $L_{\mathbf{X}}$ has property (ii) in Lemma 12.6. By switching the notation for u and v and X and Y if necessary, we may assume that X is a white tile. Since the flag $(\{L(u)\}, L(e), L(X))$ is positively-oriented, and $\phi(X)$ is white, it follows that for some $l \in \mathbb{Z}_k$ we have $\psi(L(u)) = l$ and $\psi(L(e)) = l$. Hence $L_{\mathbf{V}}(u) = l$ and $L_{\mathbf{E}}(e) = l$. Similarly, using that $L(Y)$ is black and that $(\{L(v)\}, L(e), L(Y))$ is positively-oriented, it follows that $L_{\mathbf{V}}(v) = l + 1$.

In other words, if we run along an oriented edge e in \mathcal{D}^1 so that a white tile lies on the left of e , then the label of the endpoints of e (given by $L_{\mathbf{V}}$) is increased by one, and decreased by one if a black tile lies on the left. Hence $L_{\mathbf{V}}$ has the property (iii) in Lemma 12.6. Moreover, we also see that the label $L_{\mathbf{E}}(e)$ is related to the labels of its endpoints as in statement (iv) of Lemma 12.6. The uniqueness of L follows. \square

Let f be a map realizing a two-tile subdivision rule $(\mathcal{D}^1, \mathcal{D}^0, L)$. We want to show that the property of f being combinatorially expanding for the Jordan curve \mathcal{C} of \mathcal{D}^0 is independent of the realization. In contrast, this is not true for expansion of the map (see Example 12.11). We require a lemma.

Lemma 12.8. *Let $f: S^2 \rightarrow S^2$ and $g: \widehat{S}^2 \rightarrow \widehat{S}^2$ be Thurston maps. Suppose that $\# \text{post}(f) \geq 3$, that $\mathcal{C} \subset S^2$ is an f -invariant Jordan curve with $\text{post}(f) \subset \mathcal{C}$, and that $h_0, h_1: S^2 \rightarrow \widehat{S}^2$ are orientation-preserving homeomorphisms satisfying $h_0|_{\text{post}(f)} = h_1|_{\text{post}(f)}$, $h_0 \circ f = g \circ h_1$, and $h_0(\mathcal{C}) = h_1(\mathcal{C})$.*

Then f is combinatorially expanding for \mathcal{C} if and only if g is combinatorially expanding for $\widehat{\mathcal{C}} := h_0(\mathcal{C}) = h_1(\mathcal{C})$.

Proof. We have $\text{post}(g) = h_0(\text{post}(f)) = h_1(\text{post}(f))$ (see the proof of (3.4)). Hence $\# \text{post}(g) = \# \text{post}(f) \geq 3$. Moreover, $\widehat{\mathcal{C}} \subset \widehat{S}^2$ is a Jordan curve with $\text{post}(g) \subset \widehat{\mathcal{C}}$. This curve is g -invariant, since

$$g(\widehat{\mathcal{C}}) = g(h_0(\mathcal{C})) = h_1(f(\mathcal{C})) \subset h_1(\mathcal{C}) = \widehat{\mathcal{C}}.$$

So the statement that g is combinatorially expanding for $\widehat{\mathcal{C}}$ is meaningful (see Definition 11.4).

Pick an orientation of \mathcal{C} . By our assumptions the map $\varphi := h_1^{-1} \circ h_0$ fixes the elements of $\text{post}(f)$ pointwise and the Jordan curve \mathcal{C} setwise. Since $\# \text{post}(f) \geq 3$ and $\text{post}(f) \subset \mathcal{C}$, this implies that φ preserves the orientation of \mathcal{C} . Since φ is an orientation-preserving homeomorphism on S^2 , the map φ sends each of the complementary components of \mathcal{C} to itself. Thus, φ is cellular for $\mathcal{D}^0 := \mathcal{D}(f, \mathcal{C})$ and we have $\varphi(c) = c$ for each cell $c \in \mathcal{D}^0$. As in the proof of Proposition 12.3, this implies that φ is isotopic to id_{S^2} rel. $\text{post}(f)$. Hence $h_0 = h_1 \circ \varphi$ is isotopic to $h_1 = h_1 \circ \text{id}_{S^2}$ rel. $\text{post}(f)$, and so there exists an isotopy $H^0: S^2 \times I \rightarrow \widehat{S}^2$ rel. $\text{post}(f)$ with $H_0^0 = h_0$ and $H_1^0 = h_1$.

As in the proof of Proposition 10.4, based Proposition 10.1 we can repeatedly lift the initial isotopy H^0 . In this way we can find isotopies $H^n: S^2 \times I \rightarrow \widehat{S}^2$ rel. $\text{post}(f)$ such that $H_t^n \circ f = g \circ H_t^{n+1}$ and $H_0^{n+1} = H_1^n$ for all $n \in \mathbb{N}_0$ and $t \in I$. Note that H^n for $n \geq 1$ is actually an isotopy rel. $f^{-1}(\text{post}(f)) \supset \text{post}(f)$.

Define homeomorphisms $h_n := H_0^n$ for $n \in \mathbb{N}_0$ (note that for $n = 0$ and $n = 1$ these maps agree with our given maps h_0 and h_1). Then $h_n \circ f = g \circ h_{n+1}$, and so

$$(12.5) \quad h_0 \circ f^n = g^n \circ h_n$$

for all $n \in \mathbb{N}_0$.

We have $h_n|_{\text{post}(f)} = h_0|_{\text{post}(f)}$ which implies

$$(12.6) \quad h_n(\text{post}(f)) = \text{post}(g)$$

for all $n \in \mathbb{N}_0$. Moreover, $h_n|_{f^{-1}(\text{post}(f))} = h_1|_{f^{-1}(\text{post}(f))}$ and so

$$(12.7) \quad h_n(f^{-1}(\text{post}(f))) = g^{-1}(\text{post}(g))$$

for $n \in \mathbb{N}_0$ as follows from (12.6) and Lemma 10.2.

Our hypotheses imply that if c is a cell in $\mathcal{D}^0(f, \mathcal{C})$, then $h_0(c)$ is a cell in $\mathcal{D}^0(g, \widehat{\mathcal{C}})$. Since the set

$$\widehat{\mathcal{D}}^n := \{h_n(c) : c \in \mathcal{D}^n(f, \mathcal{C})\}$$

is a cell decomposition of \widehat{S}^2 , it follows from this and (12.5) that g^n is cellular for $(\widehat{\mathcal{D}}^n, \mathcal{D}^0(g, \widehat{\mathcal{C}}))$. Since g^n is also cellular for the pair $(\mathcal{D}^n(g, \mathcal{C}), \mathcal{D}^0(g, \widehat{\mathcal{C}}))$, the uniqueness statement in Lemma 5.4 implies that $\widehat{\mathcal{D}}^n = \mathcal{D}^n(g, \mathcal{C})$ for all $n \in \mathbb{N}_0$. In other words, the n -cells for $(g, \widehat{\mathcal{C}})$ are precisely the images of the n -cells for (f, \mathcal{C}) under the homeomorphism h_n .

We also have

$$(12.8) \quad h_n(\mathcal{C}) = \widehat{\mathcal{C}}$$

for each $n \in \mathbb{N}_0$. This can be seen by induction on n as follows. The statement is true for $n = 0$ and $n = 1$ by our hypothesis and by the definition of $\widehat{\mathcal{C}}$. Assume that $h_n(\mathcal{C}) = \widehat{\mathcal{C}}$ for some $n \in \mathbb{N}$. Then by Lemma 10.2 and the induction hypotheses we have

$$J := h_{n+1}(\mathcal{C}) \subset h_{n+1}(f^{-1}(\mathcal{C})) = g^{-1}(h_n(\mathcal{C})) = g^{-1}(\widehat{\mathcal{C}}).$$

It follows from (12.7) that $(H_t^n) \circ h_n^{-1}$ is an isotopy on \widehat{S}^2 rel. $g^{-1}(\text{post}(g))$. It isotopes $\widehat{\mathcal{C}} = h_n(\mathcal{C}) \subset g^{-1}(\widehat{\mathcal{C}})$ into $J = h_{n+1}(\mathcal{C})$ rel. $g^{-1}(\text{post}(g))$. So \mathcal{C} and J are Jordan curves contained in the 1-skeleton $g^{-1}(\widehat{\mathcal{C}})$ of $\mathcal{D}^1(g, \widehat{\mathcal{C}})$ that are isotopic relative to the set $g^{-1}(\text{post}(g))$ of vertices of $\mathcal{D}^1(g, \widehat{\mathcal{C}})$. Lemma 10.12 implies that $J = \widehat{\mathcal{C}}$, and (12.8) follows.

Now (12.8) and (12.6) imply that a chain of n -tiles for (f, \mathcal{C}) joins opposite sides of \mathcal{C} if and only if their images under h_n form a chain joining opposite sides of $\widehat{\mathcal{C}}$. Since the images of the n -tiles for (f, \mathcal{C})

under h_n are the precisely the n -tiles for $(g, \widehat{\mathcal{C}})$, we have $D_n(f, \mathcal{C}) = D_n(g, \widehat{\mathcal{C}})$ for each $n \in \mathbb{N}_0$. The statement follows. \square

Now we can show the desired realization independence of combinatorial expansion.

Lemma 12.9. *Let $(\mathcal{D}^1, \mathcal{D}^0, L)$ be a two-tile subdivision rule on S^2 and \mathcal{C} be the Jordan curve of \mathcal{D}^0 . Suppose that the maps $f: S^2 \rightarrow S^2$ and $g: S^2 \rightarrow S^2$ both realize the subdivision rule and that $\# \text{post}(f) = \# \text{post}(g) \geq 3$. Then f is combinatorially expanding for \mathcal{C} if and only if g is combinatorially expanding for \mathcal{C} .*

Proof. Let \mathbf{V}^0 and \mathbf{V}^1 be the set of vertices of \mathcal{D}^0 and \mathcal{D}^1 , respectively. Then $P := \text{post}(f) = \text{post}(g) \subset \mathbf{V}^0 \subset \mathbf{V}^1$.

It follows from the proof of the uniqueness part of Proposition 12.3 that there exists a homeomorphism $h_1: S^2 \rightarrow S^2$ isotopic to id_{S^2} rel. $\mathbf{V}^1 \supset \text{post}(f) = \text{post}(g)$ that satisfies $f = g \circ h_1$. Moreover, $h_1(e) = e$ for each edge e in \mathcal{D}^1 . Since \mathcal{D}^1 is a refinement of \mathcal{D}^0 and so the 1-skeleton \mathcal{C} of \mathcal{D}^0 is contained in the 1-skeleton of \mathcal{D}^1 , this implies $h_1(\mathcal{C}) = \mathcal{C}$. Define $h_0 = \text{id}_{S^2}$. Since h_1 is isotopic to id_{S^2} rel. P we have $h_1|_P = \text{id}_{S^2}|_P = h_0|_P$. Moreover, $h_0 \circ f = g \circ h_1$, $h_1(\mathcal{C}) = \mathcal{C} = h_0(\mathcal{C})$, and both h_0 and h_1 are orientation-preserving homeomorphisms on S^2 . This shows that the hypotheses of Lemma 12.8 are satisfied (with $\widehat{S}^2 = S^2$), and so f is combinatorially expanding for \mathcal{C} if and only if g is combinatorially for $\widehat{\mathcal{C}} = h_0(\mathcal{C}) = h_1(\mathcal{C}) = \mathcal{C}$. \square

Based on the previous lemma we say that a two-tile subdivision rule $(\mathcal{D}^1, \mathcal{D}^0, L)$ is *combinatorially expanding* if one (and hence each) map f that realizes the subdivision rule is combinatorially expanding for the Jordan curve \mathcal{C} of \mathcal{D}^0 ; here we tacitly assume that the hypothesis $\# \text{post}(f) \geq 3$ of the previous lemma is true.

12.1. Examples of two-tile subdivision rules. We present some examples of two-tile subdivision rules. A first example can be obtained from the map g in Section 1.2 and the subdivision rule as indicated in Figure 1.

Example 12.10. Our next example is as follows. The white 0-tile is the (closure of the) upper half-plane, the black 0-tile is the (closure of the) lower half-plane in $\widehat{\mathbb{C}}$. The 0-vertices are the points $-1, 0, \infty$. Thus the 0-edges are $[-\infty, -1], [-1, 0], [0, \infty]$. The cell decomposition \mathcal{D}^0 is indicated to the right of Figure 5.

The white 1-tiles are the first and third quadrant, the black 1-tiles are the second and forth quadrant. The 1-vertices and their labelings

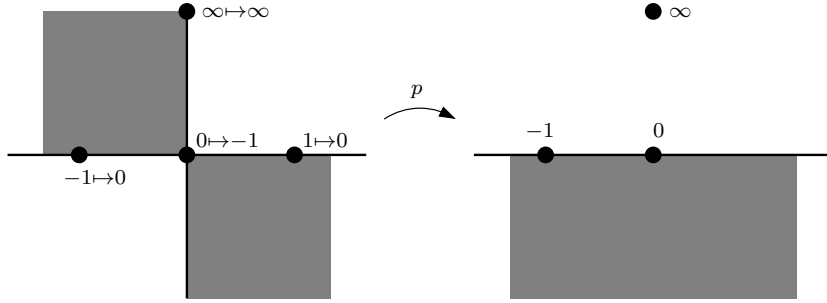
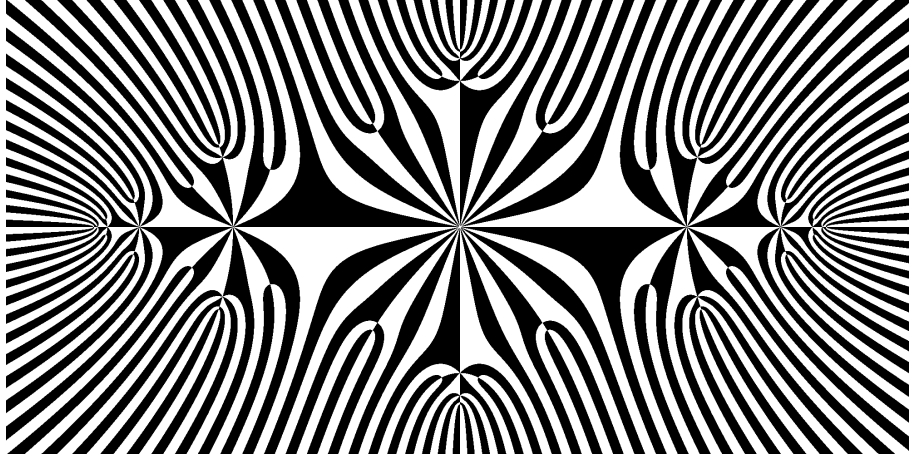
FIGURE 5. The two-tile subdivision rule for $z^2 - 1$.

FIGURE 6. Tiles of order 7 for Example 12.10.

are as follows. The point ∞ is the only 1-vertex labeled ∞ , the 1-vertices $-1, 1$ are labeled 0 , the 1-vertex 0 is labeled -1 . The cell decomposition is indicated to the left in Figure 5.

Here and in the following a point that is marked “ $a \mapsto b$ ” is a 1-vertex a that is also a 0-vertex and that is labeled by the 0-vertex b . Thus the map realizing the two-tile subdivision rule will map a to b . Similarly “ $\mapsto b$ ” marks a 1-vertex that is labeled by the 0-vertex b ; thus the realizing map will map this 1-vertex to b .

The pair $(\mathcal{D}^1, \mathcal{D}^0)$ together with this orientation-preserving labeling L is a two-tile subdivision rule. It is straightforward to check that we can choose the map $p: S^2 \rightarrow S^2$ that is generated by $(\mathcal{D}^1, \mathcal{D}^0, L)$ according to Proposition 12.5 as the map $f_1(z) = z^2 - 1$.

This two-tile subdivision rule is not combinatorially expanding. Namely, the point ∞ is the only preimage of itself by p . Thus every n -tile contains ∞ . Since the n -tiles cover the whole sphere, there has to be an

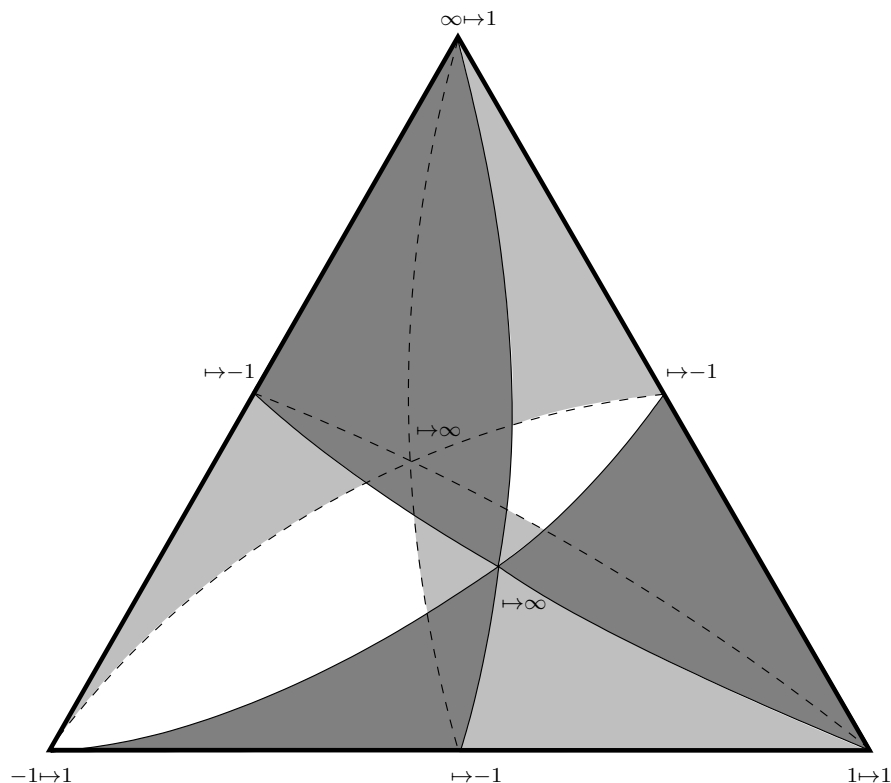


FIGURE 7. The barycentric subdivision rule.

n -tile containing both $0, \infty$. This shows that the subdivision rule is not combinatorially expanding. In fact, for all n there exist two n -tiles that contain all postcritical points $-1, 0, \infty$. Figure 6 shows the tiles of order 7.

Example 12.11 (The barycentric subdivision rule). We glue two equilateral triangles together along their boundaries to form a polyhedral surface S^2 that is conformally equivalent to $\hat{\mathbb{C}}$. The two triangles are the 0-tiles. We can find a conformal equivalence of S^2 with $\hat{\mathbb{C}}$ such that the triangles correspond to the upper and lower half-planes, and the vertices to the points $-1, 1, \infty$. For convenience we identify the vertices with $-1, 1, \infty$; they are the 0-vertices. The 0-edges are the three edges of the triangles. The bisectors divide each triangle (each 0-tile) into 6 smaller triangles. These 12 small triangles are the 1-tiles. The labeling of the 1-vertices is indicated in Figure 7. Again we obtain a two-tile subdivision rule. We can realize this subdivision rule by a map f_2 that *conformally* maps 1-tiles to the 0-tiles. Under the indicated

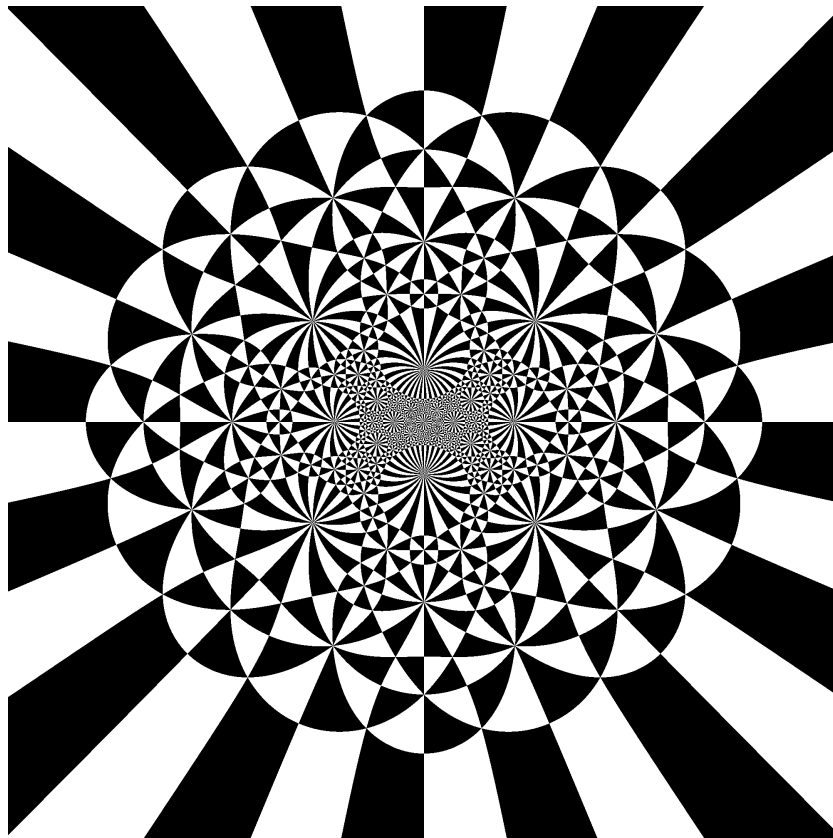


FIGURE 8. Tiles of order 4 for the barycentric subdivision rule.

identification of S^2 with $\widehat{\mathbb{C}}$, the map is then given by

$$f_2(z) = 1 - \frac{54(z^2 - 1)^2}{(z^2 + 3)^3},$$

see [CFKP, Example 4.6]. The subdivision rule is combinatorially expanding. The map f_2 however is not expanding. This follows from Proposition 19.1 as the point 1 is both a critical and a fixed point of r . One can show that the Julia set of f_2 is a Sierpiński carpet, i.e., a set homeomorphic to the standard Sierpiński carpet.

It is possible however to choose a different realization of the two-tile subdivision rule with the given labeling as in Figure 7 by a map \tilde{f}_2 that is *expanding*. Namely, we use *affine* maps to map the 1-tiles (the small triangles in the barycentric subdivision rule of the equilateral triangles) to the 0-tiles. In this case the n -tiles are Euclidean triangles for each $n \in \mathbb{N}$. The collection of all n -tiles is obtained from the $(n - 1)$ -tiles similarly as the 1-tiles were constructed from the 0-tiles: one

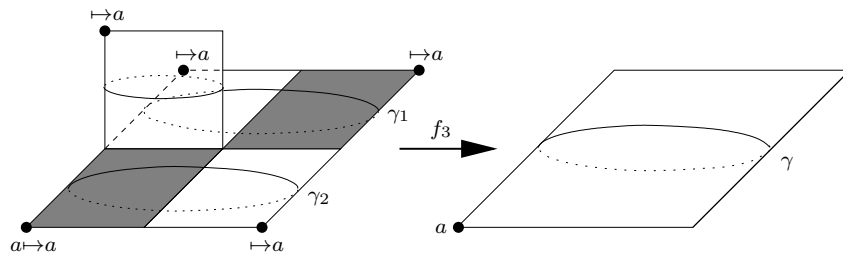


FIGURE 9. The subdivision rule for Example 12.12.

subdivides each Euclidean triangle representing an $(n - 1)$ -tile by its bisectors. It is clear that the diameters of n -tiles tend to 0 as $n \rightarrow \infty$. Hence \tilde{f}_2 is expanding, and so this map is an example of an expanding Thurston map with periodic critical points.

Example 12.12. The next example is given by a subdivision rule similar to the one that was realized by the map g from Section 1.2; see Figure 1. Again we start with a sphere that is obtained by glueing together two squares (these are the two 0-tiles) along their boundary. The four vertices are the 0-vertices that divide the common boundary into four 0-edges. Each of the two squares, or more precisely 0-tiles, is divided into four squares of half the side-length. These 8 smaller squares are 1-tiles. The edges of these squares are 1-edges. We slit the sphere along one such 1-edge (say in the white 0-tile) and glue in two small squares at the slit, as indicated to the left in Figure 9. In this way we obtain two additional 1-tiles. Topologically we have subdivided the white 0-tile into six 1-tiles; the black 0-tile is subdivided into four 1-tiles. The labeling (meaning the coloring) of the 1-tiles can be seen from Figure 9. Here only the 1-vertices that are labeled by one specific 0-vertex have been indicated to keep the picture simple and avoid unnecessary detail.

Again we obtain a two-tile subdivision rule. It is realized by a Thurston map f_3 that is combinatorially expanding. It is not equivalent to a rational map, since f_3 has a *Thurston obstruction*. In the present case, where $\# \text{post}(f_3) = 4$ and f_3 has a *hyperbolic orbifold* (see for example [Mi, Appendix E] or [McM, Appendix A]) a Thurston obstruction is given by a Jordan curve $\gamma \subset S^2 \setminus \text{post}(f_3)$ with the following properties:

- The Jordan curve γ is *non-peripheral*, i.e., each component of $S^2 \setminus \gamma$ contains two postcritical points.
- Each non-peripheral component γ_j of $f_3^{-1}(\gamma)$ is homotopic to γ in $S^2 \setminus \text{post}(f_3)$.

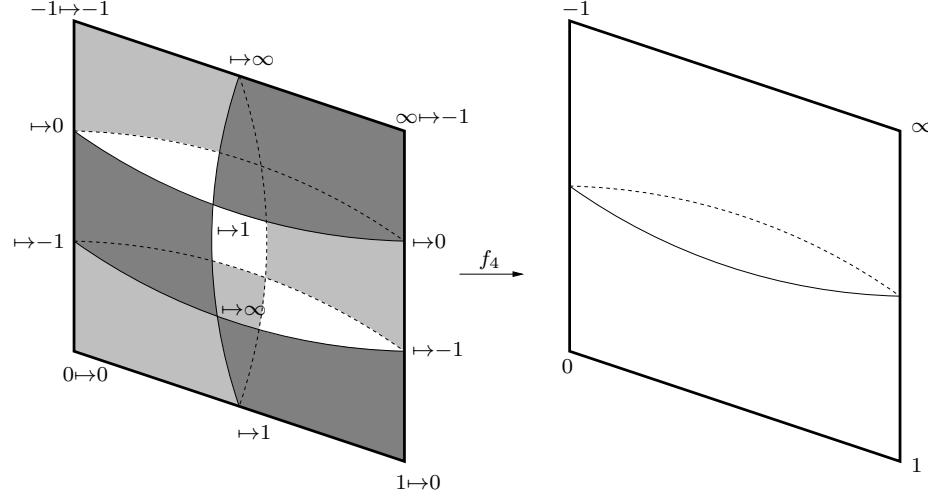


FIGURE 10. The 2-by-3 subdivision rule.

- If d_j is the degree of the map $f_3: \gamma_j \rightarrow \gamma$, then we have

$$\sum_j \frac{1}{d_j} \geq 1.$$

Thurston's theorem implies that a Thurston map f with hyperbolic orbifold and $\# \text{post}(f) = 4$ is equivalent to a rational map if and only if it has no Thurston obstruction (see [DH]). Figure 9 shows an obstruction for the map f_3 .

Example 12.13 (The 2-by-3 subdivision rule). We present another example of an expanding Thurston map f_4 that is not (Thurston) equivalent to a rational map. In a sense, this is the easiest example. However, it has a *parabolic orbifold*. Thus Thurston's criterion, as explained in the last example, does not apply.

The map f_4 will be a *Lattès-type map* (see [Yi] for precise definitions) and can be constructed in the same fashion as the Lattès map g from Section 1.2; namely, define $\psi: \mathbb{C} \rightarrow \mathbb{C}$ by setting $\psi(x + yi) = 2x + 3yi$ for $x, y \in \mathbb{R}$. Then $f_4: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ is the unique map that makes the diagram (1.2) commutative (with g replaced by f_4). This map is a realization of the two-tile subdivision rule shown in Figure 10. We form a pillow P by glueing two unit squares together along their boundaries. The two squares, as well as the four common sides and vertices of the squares, form the 0-tiles, 0-edges, 0-vertices, respectively.

Each of the two faces (i.e., squares) of the pillow is divided into 6 rectangles as shown in the figure. These 12 rectangles are the 1-tiles. Their sides and vertices are the 1-edges and 1-vertices. The coloring of 1-tiles, as well as the labeling of the 1-vertices, is indicated on the left

of Figure 10. The map f_4 sends each of the 12 rectangles affinely to one of the two squares forming the faces of the pillow. This implies that each n -tile is a rectangle with side lengths $1/2^n$ and $1/3^n$. In particular, f_4 is an expanding Thurston map.

The fact that f_4 is not equivalent to a rational map is well-known (see [DH, Prop. 9.7]). An argument for this fitting into the framework of our present work can be sketched as follows. In our outline we will rely on some results and concepts that we will be discussed later on.

To reach a contradiction, suppose that f_4 is equivalent to a rational map $R: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$. Then R is a Thurston map with no periodic critical points and is hence expanding (Proposition 19.1). So by Theorem 10.4 the maps f_4 and R are topologically conjugate. This in turns implies by Theorem 1.9 that if our pillow P is equipped with a visual metric d for f_4 , then (P, d) is quasisymmetrically equivalent to the standard 2-sphere (i.e., $\widehat{\mathbb{C}}$ equipped with the chordal metric). In particular, if X^0 is a 0-tile (i.e., one of the faces of the pillow P) equipped with a visual metric d , then it can be mapped into the standard 2-sphere by a quasisymmetric map.

Now there are visual metrics for f_4 with expansion factor $\Lambda = 2$ (it is not hard to see this directly; it also follows from the general argument in the proof of Theorem 1.7 presented in Section 18; indeed, if \mathcal{C} is the boundary of the pillow (which is f_4 -invariant), then we have $D_1 = D_1(f_4, \mathcal{C}) = 2$ in (18.4)). If d is such a metric, then (X^0, d) is bi-Lipschitz equivalent to a Rickman's rug R_α . Here by definition the Rickman's rug R_α for $0 < \alpha < 1$ is the unit square $[0, 1]^2 \subset \mathbb{R}^2$ equipped with the metric d_α given by

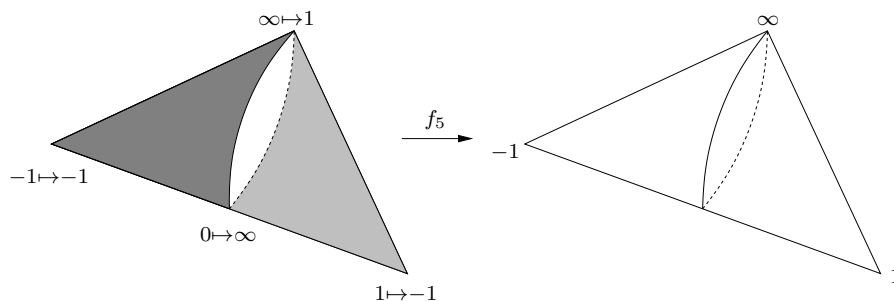
$$d_\alpha((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|^\alpha$$

for $(x_1, y_1), (x_2, y_2) \in [0, 1]^2$. In our case, (X^0, d) is bi-Lipschitz equivalent to R_α with $\alpha = \log 2 / \log 3$. It is well-known that no quasisymmetric map can lower the Hausdorff dimension

$$\dim_H(R_\alpha) = 1 + \log 3 / \log 2 > 2$$

of R_α (see [He, Theorem 15.10]); in particular, R_α and hence also (X^0, d) , cannot be mapped into the standard 2-sphere by a quasisymmetric map. This is a contradiction showing that f_4 is not Thurston equivalent to a rational map.

Example 12.14. We conclude this section by giving a whole class of examples that are similar to Example 12.12. The map f_6 defined below will be used to illustrate the construction of the invariant curve in Example 15.1, see Figure 18.

FIGURE 11. The subdivision rule realized by $1 - 2/z^2$.

The simplest two-tile subdivision of this class is given as follows. Consider a right-angled, isosceles Euclidean triangle T (thus its angles are $\pi/2, \pi/4, \pi/4$). The perpendicular bisector of the hypotenuse divides T into two triangles similar to T (scaled by the factor $\sqrt{2}$). Glue two copies of the triangle T together along their boundaries to form a pillow (i.e., a topological sphere) as before. The two faces of the pillow (i.e., the two copies of T) are the 0-tiles, and the common sides and vertices of these faces are the 0-edges and 0-vertices.

We divide each of the 0-tiles (i.e., each face of the pillow) along the perpendicular bisector of the hypotenuse. The four triangles thus obtained are the 1-tiles. Their vertices and sides are the 1-vertices and 1-edges. If the labeling is as indicated on the right of Figure 11, then we obtain a two-tile subdivision rule that can be realized by the map $f_5(z) = 1 - 2/z^2$ (this is actually a Lattès map).

As in Example 12.12, we can “add a flap” to modify the subdivision rule. More precisely, we cut the pillow along the 1-edge that is the perpendicular bisector of the hypotenuse of the white 0-tile. We take a copy of the pillow, scale it by the factor $1/\sqrt{2}$, and cut it along one leg. We then glue the two sides of the slit to corresponding sides of the slit on the original pillow. This is indicated on the left in Figure 12, where we also show the coloring of 1-tiles and the labeling of the 1-vertices.

We also show the same subdivision rule in Figure 13. Here the two triangles drawn with a thick line are the 0-tiles, the white is drawn to the right, the black to the left. Their edges are the 0-edges, their vertices the 0-vertices. To obtain a topological sphere we have to match the two pairs of 0-edges with the same markings.

The white 0-tile is subdivided into four 1-tiles, the black 0-tile into two 1-tiles. The labeling of the 1-vertices is shown in Figure 14. As before this yields a two-tile subdivision rule. It can be realized by the

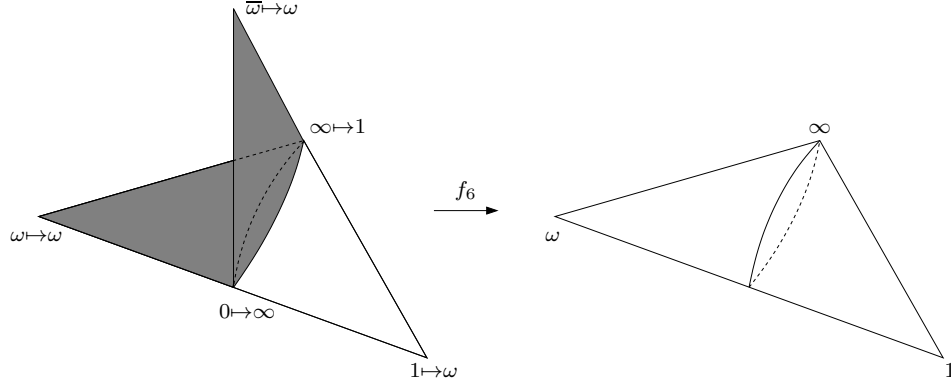
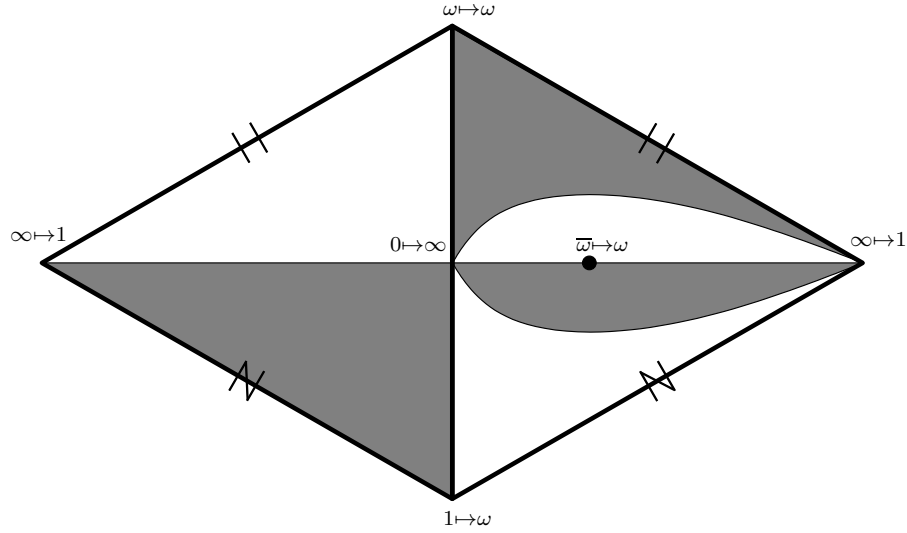


FIGURE 12. Adding a flap.

FIGURE 13. The subdivision rule realized by f_6 .

rational map f_6 given by

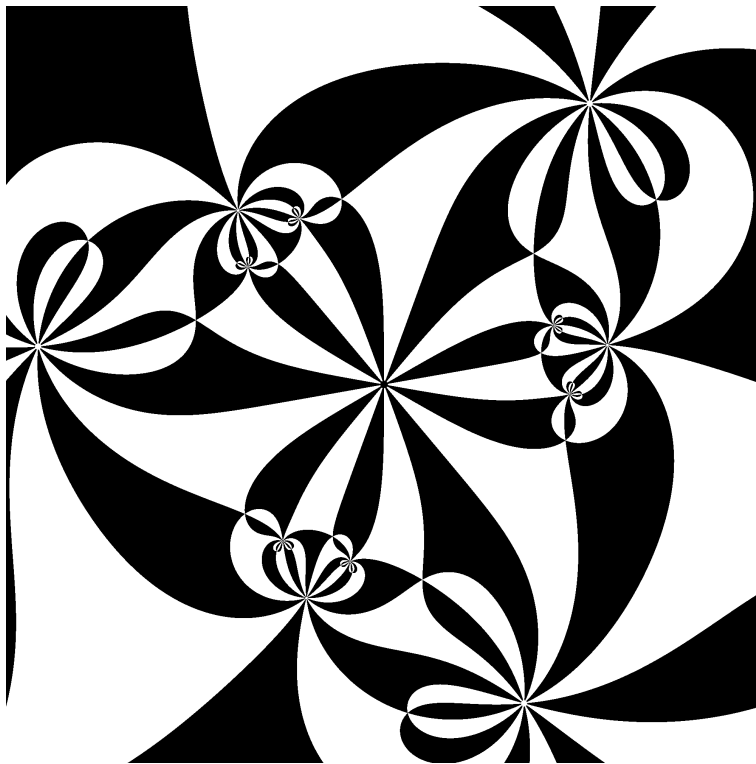
$$f_6(z) = 1 + \frac{\omega - 1}{z^3},$$

where $\omega = e^{4\pi i/3}$.

The previous example can be modified as follows. Instead of adding one flap to the 1-edge bisecting the white 0-tile in Figure 11, we can add n flaps. Similarly we can glue in m flaps at the 1-edge bisecting the black 0-tile. The resulting subdivision rule can be realized by the rational map f_7 given by

$$f_7(z) = 1 + \frac{\omega - 1}{z^d},$$

where $d = n + m + 2$ and $\omega = e^{2\pi i \frac{n+1}{d}}$.

FIGURE 14. Tiles of order 4 of the map f_6 .

The map f_6 will be used to illustrate the construction of an invariant Jordan curve \mathcal{C} with $\text{post}(f_6) \subset \mathcal{C}$. The tiles of order 4 are shown in Figure 14.

More examples can be found in [CFKP] and [Me02]. In [Me02] and [Me09a] two-tile subdivision rules realizable by rational maps were used to show that certain self-similar surfaces are *quasispheres*, i.e., quasymmetric images of unit sphere in \mathbb{R}^3 . More general examples of subdivisions can be found in [CFP06b].

The Figures 6, 8, and 14 show *symmetric conformal tilings*. This means that if two tiles share an edge, they are conformal reflections of each other along this edge. Then the tiling can be obtained by successive reflections, and so each tile encodes the information for the whole tiling.

13. COMBINATORIALLY EXPANDING THURSTON MAPS

The purpose of this section is to establish the following fact.

Proposition 13.1. *Let $F: S^2 \rightarrow S^2$ be a Thurston map that has an invariant Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(F) \subset \mathcal{C}$. If F is combinatorially expanding for \mathcal{C} , then there exists an expanding Thurston map $\tilde{F}: \tilde{S}^2 \rightarrow \tilde{S}^2$ that is Thurston equivalent to F and has an \tilde{F} -invariant Jordan curve $\tilde{\mathcal{C}} \subset \tilde{S}^2$ with $\text{post}(\tilde{F}) \subset \tilde{\mathcal{C}}$. Moreover, there exist orientation-preserving homeomorphisms $h_0, h_1: S^2 \rightarrow \tilde{S}^2$ that are isotopic rel. $\text{post}(F)$ and satisfy $h_0 \circ F = \tilde{F} \circ h_1$ and $h_0(\mathcal{C}) = \tilde{\mathcal{C}} = h_1(\mathcal{C})$.*

So up to Thurston equivalence every *combinatorially expanding* Thurston map with an invariant Jordan curve can be promoted to an *expanding* Thurston map with an invariant curve.

The proof will occupy the rest of the section. The idea is to introduce a suitable equivalence relation \sim on the sphere S^2 on which F acts and use Moore's Theorem (see Theorem 13.4) to show that the quotient space S^2/\sim is also a 2-sphere. The map \tilde{F} will then be the induced map on S^2/\sim .

So let $F: S^2 \rightarrow S^2$ be a Thurston map as in the proposition. We fix an invariant F -Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(F) \subset \mathcal{C}$ for which F is combinatorially expanding. As before we denote by \mathbf{X}^n , \mathbf{E}^n and \mathbf{V}^n the set of n -tiles, n -edges, and n -vertices, respectively, of the cell decomposition $\mathcal{D}^n = \mathcal{D}^n(F, \mathcal{C})$ defined in Section 6. A subset $\tau \subset S^2$ is called a *tile* if it is an n -tile for some $n \in \mathbb{N}_0$. We use the terms *edge*, *vertex*, *cell* in a similar way. In particular, in this section the term “cell” will always be used with this specific meaning. We will use the term *topological cell* to refer to the more general notion of cells as defined in Section 4.

Since \mathcal{C} is F -invariant, \mathcal{D}^{n+k} is a refinement of \mathcal{D}^n for all $n, k \in \mathbb{N}_0$. For each $X \in \mathbf{X}^{n+k}$ there exists a unique $Y \in \mathbf{X}^n$ with $X \subset Y$. Conversely, each n -tile Y is equal to the union of all $(n+k)$ -tiles contained in Y , and similarly each n -edge e is equal to the union of all $(n+k)$ -edges contained in e (all this was proved in Proposition 11.1). We will use this fact that cells are subdivided by cells of the same dimension and higher order repeatedly in the following.

As in the beginning of Section 11 we denote by $\mathcal{S} = \mathcal{S}(F, \mathcal{C})$ the set of all sequences $\{X^n\}$ with $X^n \in \mathbf{X}^n$ for $n \in \mathbb{N}_0$ and

$$X^0 \supset X^1 \supset X^2 \supset \dots$$

We know (see Lemma 11.2) that expansion of a Thurston map with an invariant curve is characterized by the condition that $\bigcap_n X^n$ is always a singleton set if $\{X^n\} \in \mathcal{S}$. This may not be that case for our given map F , and so we want to identify all points in such an intersection $\bigcap_n X^n$. This will not lead to an equivalence relation, since transitivity

may fail. As we will see, this issue is resolved if we define the relation as follows.

Definition 13.2. Let $x, y \in S^2$ be arbitrary. We write $x \sim y$ if and only if for all $\{X^n\}, \{Y^n\} \in \mathcal{S}$ with $x \in \bigcap_n X^n$ and $y \in \bigcap_n Y^n$ we have $X^n \cap Y^n \neq \emptyset$ for all $n \in \mathbb{N}_0$.

Recall from (7.4) that $D_n = D_n(F, \mathcal{C})$ denotes the minimal number of n -tiles forming a connected set K^n joining opposite sides of \mathcal{C} . Since F is combinatorially expanding for \mathcal{C} (see Definition 11.4), we have $\# \text{post}(F) \geq 3$ and so the term “joining opposite sides” is meaningful (see Definition 7.6). Moreover, there exists $n_0 \in \mathbb{N}$ such that $D_{n_0}(F, \mathcal{C}) \geq 2$, and so by Lemma 11.3 we have $D_n = D_n(F, \mathcal{C}) \rightarrow \infty$ as $n \rightarrow \infty$. In combination with Lemma 7.10 this implies that if τ, σ are disjoint k -cells and K^n is a connected set of n -tiles with $\sigma \cap K^n \neq \emptyset$ and $\tau \cap K^n \neq \emptyset$, then the number of tiles in K^n tends to infinity and so cannot stay bounded as $n \rightarrow \infty$. We will use this fact in the proof of the following lemma.

Lemma 13.3. *The relation \sim is an equivalence relation on S^2 .*

Proof. Reflexivity and symmetry of the relation \sim are clear. To show transitivity, let $x, y, z \in S^2$ be arbitrary and assume that $x \sim y$ and $y \sim z$. Let $\{X^n\}, \{Z^n\} \in \mathcal{S}$ with $x \in \bigcap_n X^n$ and $z \in \bigcap_n Z^n$ be arbitrary. We have to show that $X^n \cap Z^n \neq \emptyset$ for all $n \in \mathbb{N}_0$.

If this is not the case, then there exists $n_0 \in \mathbb{N}_0$ such that $X^{n_0} \cap Z^{n_0} = \emptyset$. To reach a contradiction, pick a sequence $\{Y^n\} \in \mathcal{S}$ with $y \in \bigcap_n Y^n$. Since $x \sim y$ and $y \sim z$, we have $X^n \cap Y^n \neq \emptyset$ and $Y^n \cap Z^n \neq \emptyset$ for all $n \in \mathbb{N}_0$. Then $X^{n_0} \cap Y^n \supset X^n \cap Y^n \neq \emptyset$ and $Z^{n_0} \cap Y^n \supset Z^n \cap Y^n \neq \emptyset$ for all $n \geq n_0$. So the n -tile Y^n connects the disjoint n_0 -tiles X^{n_0} and Z^{n_0} for all $n \geq n_0$. As we discussed, this is impossible by Lemma 7.10. \square

It is clear that \sim is the “smallest” equivalence relation such that all points in an intersection $\bigcap_n X_n$ with $\{X^n\} \in \mathcal{S}$ are equivalent.

If $x \in S^2$ we denote by $[x] \subset S^2$ the equivalence class of x with respect to the equivalence relation \sim , and by

$$\tilde{S}^2 = S^2 / \sim = \{[x] : x \in S^2\}$$

the quotient space of S^2 under \sim . So \tilde{S}^2 consists of all equivalence classes of \sim . Such an equivalence class is both a point in \tilde{S}^2 and a subset of S^2 . We equip \tilde{S}^2 with the quotient topology. Then the quotient map $\pi: S^2 \rightarrow \tilde{S}^2$, $x \in S^2 \mapsto [x]$, is continuous.

The quotient space \tilde{S}^2 is a topological 2-sphere. Our first goal is to show that \tilde{S} is a topological 2-sphere. For the moment consider

an arbitrary equivalence relation \sim on the sphere S^2 . We call it *closed* if $\{(x, y) \in S^2 \times S^2 : x \sim y\}$ is a closed subset of $S^2 \times S^2$ (in the older literature the term “*upper semicontinuous*” is often used instead). This is equivalent to the following condition: If $\{x_n\}$ and $\{y_n\}$ are arbitrary convergent sequences in S^2 with $x_n \rightarrow x$ and $y_n \rightarrow y$ as $n \rightarrow \infty$, and $x_n \sim y_n$ for all $n \in \mathbb{N}$, then $x \sim y$. The equivalence classes of a closed equivalence relation on S^2 are closed and hence compact subsets of S^2 .

We need the following key theorem (see [Moo] for the original proof, [Da, p. 187, Thm. 1] for a stronger statement, and [Ca, Supplement 1] for a general discussion on the 2-sphere recognition problem).

Theorem 13.4 (Moore 1925). *Let \sim be an equivalence relation on a 2-sphere S^2 . Suppose that*

- (i) *the equivalence relation \sim is closed,*
- (ii) *each equivalence class of \sim is a connected subset of S^2 ,*
- (iii) *the complement of each equivalence class of \sim is a connected subset of S^2 ,*
- (iv) *there are at least two distinct equivalence classes.*

Then the quotient space S^2/\sim is homeomorphic to S^2 .

Here it is understood that S^2/\sim is equipped with the quotient topology.

To apply this theorem in our situation, we need some preparation.

Lemma 13.5. *Let $x, y \in S^2$ be arbitrary. Then the following conditions are equivalent:*

- (i) $x \sim y$,
- (ii) *there exist $\{X^n\}, \{Y^n\} \in \mathcal{S}$ with $x \in \bigcap_n X^n$, $y \in \bigcap_n Y^n$, and $X^n \cap Y^n \neq \emptyset$ for all $n \in \mathbb{N}_0$,*
- (iii) *for all cells $\sigma, \tau \subset S^2$ with $x \in \sigma$, $y \in \tau$, we have $\sigma \cap \tau \neq \emptyset$.*

Proof. The implication (i) \Rightarrow (ii) is clear.

To show the reverse implication (ii) \Rightarrow (i), we assume that there exist $\{X^n\}, \{Y^n\} \in \mathcal{S}$ with $x \in \bigcap_n X^n$, $y \in \bigcap_n Y^n$, and $X^n \cap Y^n \neq \emptyset$ for all $n \in \mathbb{N}_0$. We claim that if $\{U^n\}, \{V^n\} \in \mathcal{S}$ are two other sequences with $x \in \bigcap_n U^n$ and $y \in \bigcap_n V^n$, then $U^n \cap V^n \neq \emptyset$ for all $n \in \mathbb{N}_0$. To reach a contradiction assume that $U^{n_0} \cap V^{n_0} = \emptyset$ for some $n_0 \in \mathbb{N}_0$. We then have

$$U^{n_0} \cap X^n \supset \{x\} \neq \emptyset \quad \text{and} \quad V^{n_0} \cap Y^n \supset \{y\} \neq \emptyset$$

for all $n \in \mathbb{N}$. Moreover, $X^n \cap Y^n \neq \emptyset$, and so for each $n \in \mathbb{N}_0$, the $K^n := X^n \cup Y^n$ is connected, consists of two n -tiles, and meets the

disjoint n_0 -tiles U^{n_0} and V^{n_0} . As before this contradicts Lemma 7.10. Hence $x \sim y$ as desired.

The implication (iii) \Rightarrow (i) is again clear. To prove (i) \Rightarrow (iii), suppose that $x \sim y$. We argue by contradiction and assume that there exist cells σ, τ with $x \in \sigma$, $y \in \tau$ and $\sigma \cap \tau = \emptyset$.

By subdividing the cells if necessary we may assume that σ and τ are cells on the same level n_0 .

There are sequences $\{X^n\}, \{Y^n\} \in \mathcal{S}$ with $x \in \bigcap_n X^n$, $y \in \bigcap_n Y^n$, $\sigma \subset X^{n_0}$ and $\tau \subset Y^{n_0}$. Since $x \sim y$, we have $X^n \cap Y^n \neq \emptyset$ for all n .

This implies that for $n \geq n_0$ the set $K^n = X^n \cup Y^n$ is connected and consists of at most two n -tiles. Moreover,

$$K^n \cap \sigma \supset X^n \cap \sigma \supset \{x\} \neq \emptyset,$$

and similarly, $K^n \cap \tau \neq \emptyset$. Hence K^n connects the disjoint n_0 -cells σ and τ . Since F is combinatorially expanding, this is impossible by Lemma 7.10 for large n . This gives the desired contradiction. \square

We now describe the geometry of the equivalence classes of \sim . This will be used in verifying the conditions in Theorem 13.4.

Lemma 13.6. *Let $M \subset S^2$ be an arbitrary equivalence class with respect to \sim . Then for each $n \in \mathbb{N}_0$ there exists a simply connected region $\Omega^n \subset S^2$ with the following properties: the set $\overline{\Omega}^n$ consists of n -tiles, we have $\Omega^{n+1} \subset \Omega^n$ for all $n \in \mathbb{N}_0$, and*

$$(13.1) \quad M = \bigcap_n \Omega^n = \bigcap_n \overline{\Omega}^n.$$

Proof. We consider three cases. In each case it is enough to define the sets Ω^n for $n \geq m$, where $m \in \mathbb{N}_0$ is suitably chosen. We put $\Omega^n = S^2$ for $n < m$. We will establish (13.1) by showing that $\bigcap_n \overline{\Omega}^n \subset M$ and $M \subset \bigcap_n \Omega^n$ in all cases.

Case 1. M contains a vertex.

In this case $M = [v]$ where v is an m -vertex for some $m \in \mathbb{N}_0$. Then v is also an n -vertex for $n \geq m$. Let $\Omega^n = W^n(v)$ for $n \geq m$ be the n -flower of v (see Definition 7.1). Then Ω^n is a simply connected region and $\overline{\Omega}^n$ consists of n -tiles as follows from Lemma 7.2 (i) and (ii). Moreover, the definition of an n -flower and Lemma 4.7 imply that $\Omega^m \supset \Omega^{m+1} \supset \Omega^{m+2} \dots$

Let $x \in \bigcap_n \overline{\Omega}^n$ be arbitrary. We want to show that $x \in M = [v]$. If this is not the case, then there exists $n_0 \geq m$ and an n_0 -tile X^{n_0} such that $x \in X^{n_0}$, but $v \notin X^{n_0}$. On the other hand, since $x \in \overline{\Omega}^{n_0}$, for each $n \geq n_0$ there exists an n -tile Y^n such that $v, x \in Y^n$. Then Y^n meets

the disjoint n_0 -cells $\{v\}$ and X^{n_0} for all $n \geq n_0$. This is impossible by Lemma 7.10. It follows that $\bigcap_n \overline{\Omega}^n \subset M$.

If $y \notin \Omega^n = W^n(v)$ for some $n \geq m$, then there exists an n -cell τ such that $v \notin \tau$ and $y \in \tau$. Since v is a vertex, this implies $v \not\sim y$ by Lemma 13.5. So $M = [v] \subset \bigcap_n \Omega^n$, and we have

$$\bigcap_n \overline{\Omega}^n \subset M \subset \bigcap_n \Omega^n,$$

showing that the three sets in this inclusion chain are the same.

We conclude that the sets Ω^n have all the desired properties.

Case 2. M does not contain a vertex, but meets an edge.

In this case we can find $m \in \mathbb{N}_0$, an m -edge e^m and a point $x \in e^m$ such that $M = [x]$. By our assumption, x is not a vertex and hence an interior point of each edge that contains x . So $x \in \text{int}(e^m)$, and, since the $(m+1)$ -edges subdivide e^m , there exists a unique $(m+1)$ -edge e^{m+1} such that $x \in \text{int}(e^{m+1}) \subset e^{m+1} \subset e^m$. Repeating this procedure, we obtain a nested sequence of n -edges e^n for $n \geq m$ such that x is an interior point of each e^n .

There exist precisely two distinct n -tiles X^n and Y^n that contain e^n on their boundaries. Define

$$\Omega^n = \text{int}(X^n) \cup \text{int}(e^n) \cup \text{int}(Y^n).$$

Then Ω^n is the union of two disjoint open Jordan regions and an open arc contained in the boundary of both regions. Hence Ω^n is simply connected. Moreover, $\overline{\Omega}^n = X^n \cup Y^n$ consists of two tiles.

An interior point of X^{m+1} close to $\text{int}(e^{m+1})$ belongs to X^m or to Y^m , since $X^m \cup Y^m$ is a neighborhood of each point in $\text{int}(e^m) \supset \text{int}(e^{m+1})$. This is only possible if $X^{m+1} \subset X^m$ or $X^{m+1} \subset Y^m$ (see the first part of the proof of Lemma 4.7). We may assume that notation is chosen so that $X^{m+1} \subset X^m$. We then must have $Y^{m+1} \subset Y^m$. Repeating this argument and switching notation for X^n and Y^n as necessary, we see that we may assume $X^m \supset X^{m+1} \supset X^{m+2} \dots$ and $Y^m \supset Y^{m+1} \supset Y^{m+2} \dots$. In particular, we have $\Omega^{n+1} \subset \Omega^n$ for all $n \geq m$.

Let $z \in \bigcap_n \overline{\Omega}^n$ be arbitrary. We want to show that $z \in M = [x]$. Suppose this is not true, and pick a sequence $\{Z^n\} \in \mathcal{S}$ with $z \in \bigcap_n Z^n$. Since $x \in \bigcap_n X^n$, $x \in \bigcap_n Y^n$ and $x \not\sim z$ by assumption, by condition (ii) in Lemma 13.5 there exists $n_0 \geq m$ such that $X^{n_0} \cap Z^{n_0} = \emptyset$ and $Y^{n_0} \cap Z^{n_0} = \emptyset$. This is absurd, since $z \in Z^{n_0} \cap \overline{\Omega}^{n_0} = Z^{n_0} \cap (X^{n_0} \cup Y^{n_0})$. We conclude that $\bigcap_n \overline{\Omega}^n \subset M$.

Suppose $z \notin \Omega^{n_0}$ for some $n_0 \geq m$. Then there exists an n_0 -tile or an n_0 -edge τ with $z \in \tau$ that does not meet the interior of e^{n_0} . Let u and v be the endpoints of e^{n_0} . Since these points are vertices, they do

not belong to M by assumption. Hence the set

$$\bigcap_n e^n \subset \bigcap_n \overline{\Omega}^n \subset M$$

does not contain u or v either. It follows that there exists $n_1 \geq n_0$ such that $u, v \notin \sigma := e^{n_1}$. Then $x \in \sigma$, $z \in \tau$, and $\sigma \cap \tau = \emptyset$, because

$$\sigma \cap \tau = e^{n_1} \cap \tau \subset (e^{n_0} \setminus \{u, v\}) \cap \tau = \emptyset.$$

This implies that $x \not\sim z$ by condition (iii) in Lemma 13.5 (iii). Therefore, $M \subset \bigcap \Omega^n$, and so the sets Ω^n have all the desired properties.

Case 3. M does not meet any edge.

Pick a point $x \in M$ and a sequence $\{X^n\} \in \mathcal{S}$ with $x \in \bigcap_n X^n$. Define $\Omega^n := \text{int}(X^n)$ for $n \in \mathbb{N}_0$. Then the sets Ω^n are nested simply connected regions, and $\overline{\Omega}^n = X^n$ is an n -tile. It follows from Lemma 13.5 (ii) that

$$\bigcap_n \overline{\Omega}^n \subset [x] = M.$$

It remains to show that $[x] \subset \bigcap_n \Omega^n$.

If this is not the case, then there exists a point $y \in S^2$ with $x \sim y$, and $n_0 \in \mathbb{N}_0$ such that $y \notin \Omega^{n_0}$. Since $\overline{\Omega}^{n_0}$ is an n_0 -tile, the boundary of Ω^{n_0} is a union of n_0 -edges. Since M does not meet any edge, we have $y \notin \overline{\Omega}^{n_0} = X^{n_0}$.

Now pick a sequence $\{Y^n\} \in \mathcal{S}$ with $y \in \bigcap_n Y^n$. Then $Z^n = X^n \cap Y^n \neq \emptyset$ for all $n \in \mathbb{N}_0$. The sets Z^n , $n \in \mathbb{N}_0$, are nonempty nested compact sets. Hence there exists a point $z \in \bigcap_n Z^n$. Then $x \sim z$ and so $z \in M$.

On the other hand, we have $y \notin X^{n_0}$ and so $X^{n_0} \neq Y^{n_0}$. Hence the intersection Z^{n_0} consists of n_0 -cells on the boundary of X^{n_0} and of Y^{n_0} , and is hence contained in a union of n_0 -edges. Since $z \in M \cap Z^{n_0}$ this means that M meets an edge contradicting our assumption. \square

Corollary 13.7. *Each equivalence class M of \sim is a compact connected set with connected complement $S^2 \setminus M$.*

Proof. Let M be an arbitrary equivalence class of \sim . By Lemma 13.6 the set M is the intersection of a nested sequence of compact and connected sets. Hence M is also compact and connected.

The complement of an open simply connected set in S^2 is connected. So Lemma 13.6 also shows that the complement $S^2 \setminus M$ of M is a union of an increasing sequence of connected sets. Hence $S^2 \setminus M$ is connected. \square

Lemma 13.8. *Let \sim be the equivalence relation on S^2 as in Definition 13.2. Then the quotient space $\tilde{S}^2 = S^2/\sim$ is homeomorphic to S^2 .*

Proof. By Lemma 13.3 our relation \sim is indeed an equivalence relation. It remains to verify the conditions (i)–(iv) in Theorem 13.4.

Condition (i): Let $\{x_n\}$ and $\{y_n\}$ be convergent sequences with $x_n \rightarrow x$ and $y_n \rightarrow y$ as $n \rightarrow \infty$, and suppose that $x_n \sim y_n$ for all $n \in \mathbb{N}$. We have to show that $x \sim y$. Suppose this is not the case. Then the equivalence classes $[x]$ and $[y]$ are disjoint. By Lemma 13.6 for sufficiently large n there exist simply connected nested regions U^n and V^n such that

$$[x] = \bigcap_n U^n = \bigcap_n \overline{U}^n \text{ and } [y] = \bigcap_n V^n = \bigcap_n \overline{V}^n.$$

Since $[x]$ and $[y]$ are disjoint, the sets \overline{U}^n and \overline{V}^n will also be disjoint for sufficiently large n , say $\overline{U}^{n_0} \cap \overline{V}^{n_0} = \emptyset$. On the other hand, since $U^{n_0} \supset [x]$ and $V^{n_0} \supset [y]$ are open, there exists $n_1 \in \mathbb{N}$ such that $x_{n_1} \in U^{n_0}$ and $y_{n_1} \in V^{n_0}$. Since \overline{U}^{n_0} and \overline{V}^{n_0} consist of n_0 -tiles and are disjoint, this means that there exist n_0 -tiles σ and τ with $x_{n_1} \in \sigma$, $y_{n_1} \in \tau$ and $\sigma \cap \tau = \emptyset$. Hence $x_{n_1} \not\sim y_{n_1}$ by Lemma 13.5. This is a contradiction. It follows that \sim is closed.

Conditions (ii)+(iii): The statements were proved in Corollary 13.7.

Condition (iv): There are at least two equivalence classes, because no two distinct vertices are equivalent by Lemma 13.5, and each post-critical point of F (there are at least 3 such points) is a vertex.

This shows that the conditions in Moore's theorem are satisfied, and so \tilde{S}^2 is indeed a sphere. \square

Quotients of cells and the induced cell decompositions on \tilde{S}^2 .

We now study what happens to our cells under the quotient map $\pi: S^2 \rightarrow \tilde{S}^2$. If $A \subset S^2$ is an arbitrary set, we denote by \tilde{A} its image under the projection map π . So $\tilde{A} = \pi(A) = \{[x] : x \in A\}$. We will see that if σ is an arbitrary cell (i.e., an n -cell for (F, \mathcal{C}) and some $n \in \mathbb{N}_0$), then $\tilde{\sigma}$ is a topological cell of the same dimension (Lemma 13.13). Moreover, the images $\tilde{\sigma}$ of the n -cells σ form a cell decomposition of \tilde{S}^2 (Lemma 13.14).

Consider an equivalence class $M \subset S^2$ of \sim . Depending on the cases in the proof of Lemma 13.6 we associate to M a nested sequence $c^m \supset c^{m+1} \supset \dots$ of n -cells c^n , where $n \geq m \in \mathbb{N}$, as follows:

If $M = [v]$, where v is an m -vertex we set $c^n := \{v\}$ for all $n \geq m$. When M does not contain a vertex, but meets an m -edge we set $c^n :=$

e^n for all $n \geq m$, where $e^m \supset e^{m+1} \supset \dots$ is the nested sequence of edges from Case 2 of the proof of Lemma 11.5. Finally, when M does not meet an edge, we set $c^n = X^n$ for all $n \geq m$, where $X^m \supset X^{m+1} \supset \dots$ is the nested sequence from Case 3 of the proof of Lemma 13.6. Note that in all cases $\bigcap c^n \subset M$.

The following lemma shows that the sequence $c^m \supset c^{m+1} \supset \dots$ can be used to give a criterion when a cell τ intersects the equivalence class M .

Lemma 13.9. *Let M be an arbitrary equivalence class with the associated nested sequence of cells $c^m \supset c^{m+1} \supset \dots$ defined as above. If τ is an arbitrary cell, then $\tau \cap M \neq \emptyset$ if and only if $c^n \subset \tau$ for all sufficiently large n .*

Proof. Case 1. $M = [v]$, where v is an m -vertex. Then $\tau \cap M \neq \emptyset$ if and only if $\tau \supset \{v\} = c^m = c^{m+1} = \dots$. This immediately follows from Lemma 13.5.

Case 2. M does not contain a vertex, but meets an edge. Then $c^m \supset c^{m+1} \supset \dots$ is the nested sequence of edges $e^m \supset e^{m+1} \supset \dots$ as defined in the proof of Lemma 13.6 in this case. Then $\tau \cap M \neq \emptyset$ if and only if $e^n \subset \tau$ for all large n .

Indeed, since $\emptyset \neq \bigcap_n e^n \subset M$, it is clear that the second condition implies the first one. Conversely, suppose $\tau \cap M \neq \emptyset$. By subdividing τ if necessary, we may assume that τ is a k -cell with $k \geq m$. Then τ has nonempty intersection with $\Omega^k = \text{int}(X^k) \cup \text{int}(e^k) \cup \text{int}(Y^k) \supset M$, where the notation is as in Case 2 of the proof of Lemma 13.6. Since the interiors of k -cells are pairwise disjoint, this is only possible if τ is an k -edge and $\tau = e^k$, or if τ is a k -tile and $\tau = X^k$ or $\tau = Y^k$. In any case $e^k \subset \tau$, and so $e^n \subset \tau$ for all large n .

Case 3. M does not meet an edge. Then $c^m \supset c^{m+1} \supset \dots$ is the nested sequence of tiles $X^m \supset X^{m+1} \supset \dots$ as defined in the proof of Lemma 13.6 in this case.

Then $\tau \cap M \neq \emptyset$ if and only if $X^n \subset \tau$ for large n . Again one direction is clear. For the other direction we can again assume that τ is a k -cell with $k \geq m$. As we have seen in the proof of Case 3 of Lemma 13.6, the set M lies in the interior of each tile X^n . So if $\tau \cap M \neq \emptyset$, then τ meets the interior of X^k . This is only possible if τ is a k -tile and $\tau = X^k$. Obviously, we then have $X^n \subset \tau$ for large n . \square

The following lemma states that if we pass to the quotient space $\tilde{S}^2 = S^2/\sim$, then we do not create “new” intersections or inclusions between cells.

Lemma 13.10. *If σ and τ are cells, then $\widetilde{\sigma} \cap \widetilde{\tau} = \widetilde{\sigma \cap \tau}$. Moreover, we have $\widetilde{\sigma} \subset \widetilde{\tau}$ if and only if $\sigma \subset \tau$.*

Proof. The inclusion $\widetilde{\sigma \cap \tau} \subset \widetilde{\sigma} \cap \widetilde{\tau}$ is trivial.

For the other inclusion let M be an arbitrary equivalence class and suppose that M (considered as a point in $\widetilde{S^2}$) is an element of the set $\widetilde{\sigma} \cap \widetilde{\tau} \subset \widetilde{S^2}$. Then M (considered as a subset of S^2) meets both cells σ and τ . By Lemma 13.9 there exist n -cells $c^n \subset M$ such that both $c^n \subset \sigma$ and $c^n \subset \tau$ for all sufficiently large n . In particular, for such n we have $M \cap \sigma \cap \tau \supset c^n \neq \emptyset$. Hence M (now again considered as a point in $\widetilde{S^2}$) lies in $\widetilde{\sigma \cap \tau}$, and we have $\widetilde{\sigma} \cap \widetilde{\tau} \subset \widetilde{\sigma \cap \tau}$ as desired.

In the second statement the implication $\sigma \subset \tau \Rightarrow \widetilde{\sigma} \subset \widetilde{\tau}$ is trivial. For the other implication assume that $\widetilde{\sigma} \subset \widetilde{\tau}$. Let k and n be the orders of σ and τ , respectively. For the moment we make the additional assumption that $k \geq n$.

By Lemma 11.5 there exists a vertex v such that $v \in \text{int}(\sigma)$ (note that this is trivial if σ is a 0-dimensional cell). Then $[v] \in \widetilde{\sigma} \subset \widetilde{\tau}$, and so $[v] \cap \tau \neq \emptyset$. Lemma 13.5 (iii) implies that $v \in \tau$ showing that

$$(13.2) \quad \text{int}(\sigma) \cap \tau \neq \emptyset.$$

Since $k \geq n$, the cell decomposition \mathcal{D}^k containing σ is a refinement of the cell decomposition \mathcal{D}^n containing τ . Therefore, as we have seen in the first part of the proof of Lemma 4.7, the relation (13.2) forces the inclusion $\sigma \subset \tau$.

If $k < n$, we subdivide σ into cells of order n . By the previous argument, τ will contain each of these cells, and so we always have $\sigma \subset \tau$ as desired. \square

Lemma 13.11. *Let $M \subset S^2$ be an equivalence class and $E \subset S^2$ be a finite union of edges. Then $E \cap M$ is connected.*

Proof. By subdividing the edges in the union representing E , we may assume that E consists of k -edges $\tau_1^k, \dots, \tau_N^k$, where k is large enough. Again we consider three cases for M using the notation of the proof of Lemma 13.6.

Case 1. $M = [v]$, where v is a vertex. Then, as we have seen in the proof of Case 1 of Lemma 13.9, $\tau_i^k \cap M \neq \emptyset$ if and only if $v \in \tau_i^k$. By passing to larger k and subdividing the edges τ_i^k if necessary, we may assume that $v \in \tau_i^k$ if and only if v is one of the two endpoints of τ_i^k . If v is an endpoint of τ_i^k , then there is a unique nested sequence of edges $\tau_i^k \supset \tau_i^{k+1} \supset \dots$ such that τ_i^n for $n \geq k$ is an n -edge with one of its endpoints equal to v .

It follows from the representation of M as in Case 1 of the proof of Lemma 13.6 that $M \cap \tau_i^k = \bigcap_n \tau_i^n \supset \{v\}$. In particular, this set is a point or an arc and hence connected. It follows that $E \cap M$ is empty or consists of a union of connected sets that share the point v . Hence $E \cap M$ is connected.

Case 2. M meets an edge, but does not contain a vertex. We use notation as in Case 2 of the proof of Lemma 13.6. We may assume that $k \geq m$, where m is as in the proof of this lemma. As we in this proof, the only k -edge that meets M is e^k . So $E \cap M = e^k \cap M$ or $E \cap M = \emptyset$, depending whether e^k is among the edges $\tau_1^k, \dots, \tau_N^k$ or not; moreover, $e^k \cap M = \bigcap_n e^n$ as follows from the representation of M as in Case 2 of the proof of Lemma 13.6. In particular, $e^k \cap M$ is a point or an arc, and hence connected. It follows that $E \cap M$ is connected.

Case 3. M does not meet an edge. Then $E \cap M = \emptyset$ is connected. \square

For the first main result of this subsection we will need a 1-dimensional version of Moore's Theorem. It can easily be derived from the topological characterization of arcs and topological circles (in equivalent form this as stated as two exercises in [Da, p. 21, Ex. 2 and 3]). We will give a simple direct proof for the convenience of the reader.

Lemma 13.12. *Let J be an arc or a topological circle, and \sim be an equivalence relation on J . Suppose that*

- (i) *each equivalence class of \sim is a compact and connected subset of J ,*
- (ii) *there are at least two distinct equivalence classes.*

Then the quotient space $\tilde{J} = J/\sim$ is an arc or a topological circle, respectively.

Proof. We may assume that J is equal to the unit interval $[0, 1]$ or the unit circle $\partial\mathbb{D} \subset \mathbb{C}$. We denote by $[u] \subset J$ the equivalence class of a point $u \in J$, and by $\pi: J \rightarrow J/\sim$ the quotient map that sends each point $u \in J$ to $[u]$ (considered as a point in J/\sim). By our assumptions, in both cases $J = [0, 1]$ and $J = \partial\mathbb{D}$ each set $[u]$ is a subarc of J (possibly degenerate). This implies that the equivalence relation \sim is closed. Indeed, suppose (u_n) and (v_n) are two sequences in J with $u_n \sim v_n$ for $n \in \mathbb{N}$, $u_n \rightarrow u \in J$ and $v_n \rightarrow v \in J$ as $n \rightarrow \infty$. Let $\alpha_n := [u_n] = [v_n]$ for $n \in \mathbb{N}_0$. If in the sequence $\alpha_1, \alpha_2, \dots$ one of the sets α_n appears infinitely often, then, by passing to a subsequence if necessary, we may assume that $\alpha := \alpha_1 = \alpha_2 = \dots$. Then (u_n) and (v_n) are sequences in α . Since α is a closed set, we conclude $u, v \in \alpha$, and so, since α is also an equivalence class, we have $u \sim v$ as desired.

If none of the sets appears α_n appears infinitely often in the sequence, then necessarily $\text{diam}(\alpha_n) \rightarrow 0$ as $n \rightarrow \infty$, as J does not carry infinitely many pairwise disjoint arcs with diameter bounded away from 0. This implies $|u_n - v_n| \rightarrow 0$ as $n \rightarrow \infty$, and so $u = v$; again we have $u \sim v$ showing that \sim is closed.

Since \sim is closed, the quotient space J/\sim is Hausdorff; this follows from the fact that the complement of the “diagonal” in $J \times J$ is an open set, and hence a neighborhood for each of the points it contains. Moreover, the space J/\sim is also compact, since it is the continuous image of J under the continuous map π . So J/\sim is a compact Hausdorff space in both cases $J = [0, 1]$ and $J = \partial\mathbb{D}$.

Now we restrict ourselves to the case $J = [0, 1]$. For $u, v \in [0, 1]$ we define $[u] < [v]$ iff $u < v$. This is well-defined, and a total order for the equivalence classes of \sim , i.e., for $u, v \in J$ we have precisely one of the cases $[u] < [v]$, $[v] < [u]$, or $[u] = [v]$.

We want to show that J/\sim is homeomorphic to $[0, 1]$. For this it suffices to find a surjective and continuous map $h: [0, 1] \rightarrow [0, 1]$ such that $h(u) = h(v)$ for $u, v \in [0, 1]$ if and only if $u \sim v$. For then h will descend to a continuous bijection \tilde{h} of the compact Hausdorff space J/\sim onto $[0, 1]$. Then \tilde{h} is a homeomorphism as desired.

We now construct h . By enumerating the sets $[u]$ for $u \in [0, 1] \cap \mathbb{Q}$ we can find a sequence of closed subintervals I_0, I_1, \dots of $[0, 1]$ with $I_0 = [0]$, $I_1 = [1]$ such that the intervals I_0, I_1, \dots are pairwise distinct equivalence classes for \sim and such that

$$D := \bigcup_n I_n$$

is dense in $[0, 1]$. Note that by our hypotheses we have $I_0 \neq I_1$. Moreover, whenever $x, y \in [0, 1]$ and $[u] < [v]$, then there exists $n \in \mathbb{N}_0$ with $[u] < I_n < [v]$; otherwise, each $q \in [x, y] \cap \mathbb{Q}$, and hence each point in $[u, v]$, would be contained in one of the sets $[u]$ and $[v]$; this is impossible since $[u]$ and $[v]$ are closed disjoint sets with non-empty intersection with $[u, v]$, and $[u, v]$ is connected. So between any two distinct equivalence classes there is always one of the sets I_n ; this and $I_0 \neq I_1$ imply that the list I_0, I_1, \dots is infinite.

We now inductively pick a number $y_n \in [0, 1]$ for each set I_n so that $I_n < I_k$ if and only if $y_n < y_k$. This is done as follows. Let $y_0 = 0$ and $y_1 = 1$, and suppose numbers y_0, \dots, y_n with the desired properties have been chosen for the sets I_0, \dots, I_n , where $n \geq 1$. For some bijection $\phi: \{0, \dots, n\} \rightarrow \{0, \dots, n\}$ with $\phi(0) = 0$ and $\phi(n) = 1$, we have

$$I_{\phi(0)} = I_0 < I_{\phi(1)} < \dots < I_{\phi(n)} = I_1.$$

Then there exists a unique number $l \in \{0, \dots, n-1\}$ such that

$$I_{\phi(l)} < I_{n+1} < I_{\phi(l+1)}.$$

Define $y_{n+1} := \frac{1}{2}(y_{\phi(l)} + y_{\phi(l+1)})$. It is clear that this gives a correspondence $I_n \leftrightarrow y_n$ for $n \in \mathbb{N}_0$ as desired.

If $I_n < I_k$, then there exists $m > \max\{n, k\}$ such that $I_n < I_m < I_k$. This and the definition of y_n imply that the set $D' := \{y_n : n \in \mathbb{N}_0\}$ is dense in $[0, 1]$; indeed, if we use the relation $<$ to successively order the values $y_0 = 0, y_1 = 1, y_2, \dots$, and $y_n < y_k$ are immediate neighbors in this ordering up to some point, then eventually another value y_m will fall in the “gap” (y_n, y_k) cutting it in half, and so all gaps between the values y_n will become small.

There exists a unique function $h: D \rightarrow D'$ such that $h|_{I_n} \equiv y_n$ for all $n \in \mathbb{N}_0$. Then h is non-increasing, and we can extend it to a non-increasing function on $[0, 1]$, also denoted h , by setting

$$h(u) := \sup\{h(t) : t \in D \cap [0, u]\} \text{ for } u \in [0, 1].$$

Then h is a non-increasing function on $[0, 1]$ with a dense image in $[0, 1]$. Hence h is continuous (at each point left-hand and right-hand limit exist and agree with the function value), and surjective.

It remains to verify that $h(u) = h(v)$ if and only if $u \sim v$. Suppose first that $u, v \in [0, 1]$ and $u \sim v$. If $u = v$, then $h(u) = h(v)$. If $u \neq v$, then $[u] = [v]$ is a non-degenerate interval, and so it must be among the intervals I_0, I_1, \dots by density of D , say $[u] = [v] = I_n$. Then $h(u) = y_n = h(v)$ by definition of h .

Conversely, suppose that $u, v \in [0, 1]$ and $u \not\sim v$. Without loss of generality $u < v$. Then $[u] < [v]$ and we can choose intervals I_n and I_k with $k, n \in \mathbb{N}_0$ such that $[u] < I_n < I_k < [v]$. Then the properties of h imply

$$h(u) \leq h|_{I_n} = y_n < h|_{I_k} = y_k \leq h(v),$$

and so $h(u) \neq h(v)$ as desired. This completes the proof that $J/\sim = [0, 1]/\sim$ is homeomorphic to $[0, 1]$, and hence an arc.

Note that the images of the endpoints of J , i.e., 0 and 1, under the quotient map π are the endpoints of the quotient arc J/\sim . Indeed, the endpoints of an arc α are precisely the points $p \in \alpha$ for which $\alpha \setminus \{p\}$ is connected. Note that $[0, 1] \setminus [0]$ is a half-open interval and hence connected. Hence

$$(J/\sim) \setminus \{\pi(0)\} = \pi([0, 1] \setminus [0])$$

is connected as the continuous image of the connected set $[0, 1] \setminus [0]$. So $\pi(0)$ is one of the endpoints of J/\sim , and a similar argument shows that $\pi(1)$ is the other endpoint.

We now turn to the case $J = \partial\mathbb{D}$. Then by our hypotheses we can pick points $u, v \in \partial\mathbb{D}$ with $u \not\sim v$. In particular, $u \neq v$. Let α and β be the two subarcs of $\partial\mathbb{D}$ with endpoints u and v . Define $p = \pi(u)$, $q = \pi(v)$, $\tilde{\alpha} = \pi(\alpha)$, $\tilde{\beta} = \pi(\beta)$. Then $p \neq q$, and our hypotheses imply that $\tilde{\alpha} \cap \tilde{\beta} = \{p, q\}$ (a connected set in $\partial\mathbb{D}$ that meets both α and β must meet u or v). By the first part of the proof (consider \sim restricted to α and β), the sets $\tilde{\alpha}$ and $\tilde{\beta}$ are arcs, and they have the endpoints p and q . So they have precisely their endpoints in common. Every compact Hausdorff space that can be represented as a union of two such arcs is homeomorphic to $\partial\mathbb{D}$. Hence $J = \partial\mathbb{D}/\sim$ is homeomorphic to $\partial\mathbb{D}$ and thus a topological circle. \square

Lemma 13.13. *Let τ be an edge or a tile. Then $\tilde{\tau}$ is an arc or a closed Jordan region, respectively. Moreover, $\partial\tilde{\tau} = \widetilde{\partial\tau}$.*

Here $\partial\tau$ (and similarly $\partial\tilde{\tau}$) refers as usual to the boundary of a cell τ as defined in Section 4. So $\partial\tau$ is the topological boundary of τ in S^2 if τ is a tile, and equal to the set consisting of the two endpoints of τ if τ is an edge. If τ is 0-dimensional cell, i.e., a singleton set consisting of a vertex, then $\partial\tau = \emptyset$, and the statement in the lemma is trivially true. So the lemma can be formulated in an equivalent form by saying that if $\tau \subset S^2$ is a cell (in one of the cell decompositions $\mathcal{D}^n(F, \mathcal{C})$), then $\tilde{\tau} \subset \tilde{S}^2$ is a cell (in the general topological sense) of the same dimension, and the boundary of $\tilde{\tau}$ is the image of the boundary of τ under the quotient map.

Proof. Suppose first that τ is an edge. Then our equivalence relation \sim on S^2 restricts to an equivalence relation on τ whose quotient space can be identified with the subset $\tilde{\tau}$ of \tilde{S}^2 . The equivalence classes on τ have the form $\tau \cap M$, where $M \subset S^2$ is an equivalence class with respect to \sim .

Each set $\tau \cap M$ is compact, as \sim is closed, and connected by Lemma 13.11. Moreover, τ meets a least two distinct equivalence classes, as its endpoints are vertices and hence not equivalent. Lemma 13.12 implies that $\tilde{\tau} \subset \tilde{S}^2$ is indeed an arc.

Let u and v be the two endpoints of τ . Then $[u] \cap \tau$ is a compact connected subset of τ containing u . Hence this set is a subarc of τ with one endpoint equal to u . This implies that the set $\tau \setminus [u]$ is connected, and so the set $\pi(\tau \setminus [u]) = \tilde{\tau} \setminus \{\pi(u)\}$ is also connected. Therefore, $\pi(u)$ is an endpoint of $\tilde{\tau}$ (a similar argument was presented in the proof of Lemma 13.12). By the same reasoning we see that $\pi(v)$ is also an endpoint of $\tilde{\tau}$. Since u and v are distinct vertices, we have $u \not\sim v$ and so $\pi(u) \neq \pi(v)$. Hence $\partial\tilde{\tau} = \{\pi(u), \pi(v)\} = \pi(\{u, v\}) = \widetilde{\partial\tau}$.

If τ is a tile, say an n -tile, then τ is a closed Jordan region whose boundary $J = \partial\tau$ is a topological circle consisting of finitely many edges. By the Schönflies Theorem we can write S^2 as a disjoint union $S^2 = \Omega_1 \cup J \cup \Omega_2$, where Ω_1 and Ω_2 are open Jordan regions bounded by J . Then τ coincides with one of the sets $\overline{\Omega}_1$ or $\overline{\Omega}_2$, say $\tau = \overline{\Omega}_1$.

The set $\tilde{J} \subset \tilde{S}^2$ is also a topological circle as follows from the fact that \sim is closed, Lemma 13.11, and Lemma 13.12. So we can also write \tilde{S}^2 as a disjoint union $\tilde{S}^2 = D_1 \cup \tilde{J} \cup D_2$, where D_1 and D_2 are open Jordan regions in \tilde{S}^2 bounded by \tilde{J} . If we take preimages under the quotient map $\pi: S^2 \rightarrow \tilde{S}^2$, we get the disjoint union $S^2 = \pi^{-1}(D_1) \cup \pi^{-1}(\tilde{J}) \cup \pi^{-1}(D_2)$. Since point preimages under π , i.e., equivalence classes, are connected, the map π is *monotone*, and hence preimages of connected sets are connected [Da, p. 18, Prop. 1]. So the sets $\pi^{-1}(D_1)$ and $\pi^{-1}(D_2)$ are open connected sets disjoint from $\pi^{-1}(\tilde{J}) \supset J$. It follows that each of the sets $\pi^{-1}(D_1)$ and $\pi^{-1}(D_2)$ is contained in one of the regions Ω_1 and Ω_2 .

These sets cannot be contained in the same region Ω_i . Indeed, if $\pi^{-1}(D_1) \cup \pi^{-1}(D_2) \subset \Omega_1$, for example, then $\Omega_2 \subset \pi^{-1}(\tilde{J})$, and so $\pi(\Omega_2) \subset \tilde{J}$. This means that every point in Ω_2 is equivalent to a point in J . This is impossible, because Ω_2 contains the interior of an n -tile, and hence a k -vertex for some $k > n$ (see Lemma 11.5 (ii)). Such a vertex is not equivalent to any point in J by Lemma 13.5.

By what we have just seen, we may assume that indices are chosen such that $\pi^{-1}(D_1) \subset \Omega_1$ and $\pi^{-1}(D_2) \subset \Omega_2$. Then $\overline{D}_1 = D_1 \cup \tilde{J} \subset \pi(\overline{\Omega}_1)$ and $\pi(\overline{\Omega}_1) \cap D_2 = \emptyset$. Hence $\tilde{\tau} = \pi(\overline{\Omega}_1) = \overline{D}_1$ is a closed Jordan region. Moreover, $\partial\tilde{\tau} = \partial D_1 = \tilde{J} = \partial\tau$. \square

The next lemma summarizes some of the results in this subsection.

Lemma 13.14. *Let $n, k \in \mathbb{N}_0$. Then we have:*

- (i) *For each $\tau \in \mathcal{D}^n$ the set $\tilde{\tau}$ is a topological cell in \tilde{S}^2 of the same dimension as τ .*
- (ii) *For $\sigma, \tau \in \mathcal{D}^n$, we have $\tilde{\sigma} = \tilde{\tau}$ if and only if $\sigma = \tau$.*
- (iii) *$\tilde{\mathcal{D}}^n := \{\tilde{\tau} : \tau \in \mathcal{D}^n\}$ is a cell decomposition of \tilde{S}^2 .*
- (iv) *The map $\tau \in \mathcal{D}^n \mapsto \tilde{\tau} \in \tilde{\mathcal{D}}^n$ is an isomorphism between the cell complexes \mathcal{D}^n and $\tilde{\mathcal{D}}^n$.*
- (v) *$\tilde{\mathcal{D}}^{n+k}$ is a refinement of $\tilde{\mathcal{D}}^n$.*

Proof. (i) This follows from Lemma 13.13.

(ii)–(iii) Let σ and τ be arbitrary n -cells, and suppose that $\text{int}(\tilde{\sigma}) \cap \text{int}(\tilde{\tau}) \neq \emptyset$. Pick a point $p \in \text{int}(\tilde{\sigma}) \cap \text{int}(\tilde{\tau})$. Then $p \in \tilde{\sigma} \cap \tilde{\tau} = \widetilde{\sigma \cap \tau}$

(see Lemma 13.10), and so there exists $x \in \sigma \cap \tau$ with $\pi(x) = p$. Then $x \in \text{int}(\sigma)$, for otherwise $x \in \partial\sigma$ and so $p = \pi(x) \in \partial\tilde{\sigma}$ (see Lemma 13.13), contradicting the choice of p . Similarly, $x \in \text{int}(\tau)$. So $x \in \text{int}(\sigma) \cap \text{int}(\tau)$ which implies that $\sigma = \tau$. Statement (ii) follows.

This shows that the topological cells $\tilde{\tau}$ for $\tau \in \mathcal{D}^n$ are all distinct, and no two have a common interior point. Moreover, there are finitely many of these cells, and they cover \tilde{S}^2 , because the cells in \mathcal{D}^n cover S^2 . Finally, for a cell $\tilde{\tau}$ we have $\partial\tilde{\tau} = \partial\tau$ (Lemma 13.13). Since $\partial\tau$ is a union of cells in \mathcal{D}^n , it follows that $\partial\tilde{\tau}$ is a union of cells in $\tilde{\mathcal{D}}^n$. This shows that $\tilde{\mathcal{D}}^n$ is a cell decomposition of \tilde{S}^2 .

(iii) By (i) and (ii) the map $\tau \in \mathcal{D}^n \mapsto \tilde{\tau} \in \tilde{\mathcal{D}}^n$ is a bijection between \mathcal{D}^n and $\tilde{\mathcal{D}}^n$ that preserves dimensions of cells. By Lemma 13.10 the map also satisfies condition (ii) in Definition 12.2. Hence it is an isomorphism between cell complexes.

(iv) This immediately follows from the definitions and the fact that \mathcal{D}^{n+k} is a refinement of \mathcal{D}^n . \square

The induced map \tilde{F} on \tilde{S}^2 . We will now show that F induces a map on the sphere \tilde{S}^2 and study the properties of this map.

Lemma 13.15. *Suppose that $x, y \in S^2$ and $x \sim y$. Then $F(x) \sim F(y)$.*

Proof. Let $x, y \in S^2$ with $x \sim y$ be arbitrary. Pick $\{X^n\}, \{Y^n\} \in \mathcal{S}$ with $x \in \bigcap X^n$ and $y \in \bigcap Y^n$. Define $U^n = F(X^{n+1})$ and $V^n = F(Y^{n+1})$ for $n \in \mathbb{N}_0$. Then U^n and V^n are n -tiles, and so $\{U^n\}, \{V^n\} \in \mathcal{S}$. Moreover, $F(x) \in \bigcap U^n$ and $F(y) \in \bigcap V^n$. Since $x \sim y$ we have $X^n \cap Y^n \neq \emptyset$ for all $n \in \mathbb{N}$. Hence

$$U^n \cap V^n = F(X^{n+1}) \cap F(Y^{n+1}) \supset F(X^{n+1} \cap Y^{n+1}) \neq \emptyset$$

for all $n \in \mathbb{N}_0$. Lemma 13.5 (ii) now implies that $F(x) \sim F(y)$ as desired. \square

By the previous lemma the map $\tilde{F}: \tilde{S}^2 \rightarrow \tilde{S}^2$ given by

$$\tilde{F}([x]) = [F(x)] \quad \text{for } x \in S^2$$

is well-defined. Then $\tilde{F} \circ \pi = \pi \circ F$, and it follows from the properties of the quotient topology that \tilde{F} is continuous.

In the following $\tilde{\mathcal{D}}^n = \{\tilde{\tau} : \tau \in \mathcal{D}^n(F, \mathcal{C})\}$ for $n \in \mathbb{N}_0$ will denote the cell decomposition of \tilde{S}^2 as provided by Lemma 13.14. As the next lemma shows, the map \tilde{F}^n has injectivity properties similar to F^n .

Lemma 13.16. *Let τ be a n -cell, $n \in \mathbb{N}$. Then \tilde{F}^n is a homeomorphism of $\tilde{\tau}$ onto $\tilde{\sigma}$, where $\sigma = F^n(\tau)$. In particular, \tilde{F}^n is cellular for $(\tilde{\mathcal{D}}^n, \tilde{\mathcal{D}}^0)$.*

Proof. Note that F^n is a homeomorphism of τ onto σ . Hence

$$\tilde{F}^n(\tilde{\tau}) = (\tilde{F}^n \circ \pi)(\tau) = (\pi \circ F^n)(\tau) = \tilde{\sigma}$$

showing that \tilde{F}^n maps $\tilde{\tau}$ onto $\tilde{\sigma}$.

So it remains to show the injectivity of \tilde{F}^n on $\tilde{\tau}$, or equivalently, that if $x, y \in \tau$ and $F^n(x) \sim F^n(y)$, then $x \sim y$. Since every vertex and every edge is contained in a tile, we may also assume that τ is an n -tile.

If $x, y \in \tau$ and $F^n(x) \sim F^n(y)$, then we can pick sequences $\{X^k\}$ and $\{Y^k\}$ in \mathcal{S} such that $X^n = Y^n = \tau$ and $x \in \bigcap_k X^k$, $y \in \bigcap_k Y^k$. Then $F^n(X^{k+n})$ and $F^n(Y^{k+n})$ are k -tiles for $k \in \mathbb{N}_0$. Moreover, the sequences $\{F^n(X^{k+n})\}$ and $\{F^n(Y^{k+n})\}$ are in \mathcal{S} , and $F^n(x) \in \bigcap_k F^n(X^{k+n})$, $F^n(y) \in \bigcap_k F^n(Y^{k+n})$. Since $F^n(x) \sim F^n(y)$, this implies that $F^n(X^{k+n}) \cap F^n(Y^{k+n}) \neq \emptyset$ for all $k \in \mathbb{N}_0$. Since $X^{k+n}, Y^{k+n} \subset \tau$ for $k \geq 0$ and $F^n|_{\tau}$ is injective, we conclude that $X^{k+n} \cap Y^{k+n} \neq \emptyset$ for $k \geq 0$. Since $X^n = Y^n = \tau$, we also have $X^k = Y^k$ for $k = 0, \dots, n-1$. Hence $X^k \cap Y^k \neq \emptyset$ for all $k \geq 0$. Lemma 13.5 (ii) then shows that $x \sim y$ as desired.

The fact that \tilde{F}^n is cellular for $(\tilde{\mathcal{D}}^n, \tilde{\mathcal{D}}^0)$ follows from the first part of the proof and the fact that F^n is cellular for $(\mathcal{D}^n, \mathcal{D}^0)$. \square

The auxiliary homeomorphisms h_0 and h_1 . We now want to define certain homeomorphisms $h_0, h_1: S^2 \rightarrow \tilde{S}^2$ that make the following diagram commutative:

$$\begin{array}{ccc} S^2 & \xrightarrow{h_1} & \tilde{S}^2 \\ F \downarrow & & \downarrow \tilde{F} \\ S^2 & \xrightarrow{h_0} & \tilde{S}^2. \end{array}$$

For the definition of h_0 recall that S^2 is the union of two 0-tiles X_1^0 and X_2^0 with common boundary \mathcal{C} . The Jordan curve \mathcal{C} is further decomposed into $k = \# \text{post}(F) \geq 3$ 0-edges and 0-vertices. If we consider the images of these cells of order 0 under the quotient map π , then by Lemma 13.14 we get a cell decomposition $\tilde{\mathcal{D}}^0$ of \tilde{S}^2 . It contains two tiles \tilde{X}_1^0 and \tilde{X}_2^0 whose common boundary is the Jordan curve $\tilde{\mathcal{C}} = \pi(\mathcal{C})$ (this follows from the second statement in Lemma 13.13), and k distinct vertices and edges on $\tilde{\mathcal{C}}$. There are no other cells in $\tilde{\mathcal{D}}^0$.

By constructing successive extensions of maps sending the i -skeleton of the cell decomposition \mathcal{D}^0 to the i -skeleton of $\tilde{\mathcal{D}}^0$ for $i = 0, 1, 2$ (as in the proof of the first part of Proposition 12.3), one sees that there exists a (non-unique) homeomorphism $h_0: S^2 \rightarrow \tilde{S}^2$ such that

$$h_0(\tau) = \tilde{\tau}$$

for all 0-cells τ . In particular, if v is a 0-vertex (i.e., a point in $\text{post}(F)$), then $h_0(v) = \pi(v)$. We orient \tilde{S}^2 so that h_0 is orientation-preserving.

Now let τ be an arbitrary 1-cell in S^2 . Then $F(\tau)$ is a 0-cell, and by Lemma 13.16 the map $\tilde{F}|_{\tilde{\tau}}$ is a homeomorphism of $\tilde{\tau}$ onto $\tilde{F}(\tau) = h_0(F(\tau))$. Hence the map

$$\varphi_\tau := (\tilde{F}|_{\tilde{\tau}})^{-1} \circ h_0 \circ (F|_\tau)$$

is well-defined and a homeomorphism from τ onto $\tilde{\tau}$. If $x \in \tau$, then $y = \varphi_\tau(x)$ is the unique point $y \in \tilde{\tau}$ with $\tilde{F}(y) = h_0(F(x))$.

If σ and τ are two 1-cells in S^2 with $\sigma \subset \tau$, then

$$\varphi_\tau|_\sigma = \varphi_\sigma.$$

Indeed, if $x \in \sigma$, then $y = \varphi_\sigma(x) \in \tilde{\sigma} \subset \tilde{\tau}$ and $\tilde{F}(y) = h_0(F(x))$. Hence $\varphi_\sigma(x) = y = \varphi_\tau(x)$ by the uniqueness property of $\varphi_\tau(x)$.

If a point $x \in S^2$ lies in two 1-cells τ and τ' , then $\varphi_\tau(x) = \varphi_{\tau'}(x)$. Indeed, there exists a unique 1-cell σ with $x \in \text{int}(\sigma)$. Then $\sigma \subset \tau \cap \tau'$ by Lemma 4.3 (ii), and so by what we have just seen we conclude

$$\varphi_\tau(x) = \varphi_\sigma(x) = \varphi_{\tau'}(x).$$

This allows us to define a map $h_1: S^2 \rightarrow \tilde{S}^2$ as follows. If $x \in S^2$, pick a 1-cell τ in S^2 with $x \in \tau$, and set

$$h_1(x) = \varphi_\tau(x).$$

Then h_1 is well-defined.

Lemma 13.17. *The map $h_1: S^2 \rightarrow \tilde{S}^2$ is an orientation-preserving homeomorphism of S^2 onto \tilde{S}^2 satisfying $h_0 \circ F = \tilde{F} \circ h_1$ and $h_1(\tau) = \tilde{\tau}$ for each cell τ in \mathcal{D}^1 .*

Proof. We have $h_1|_\tau = \varphi_\tau$ for each 1-cell τ . Hence h_1 is continuous if restricted to an arbitrary 1-cell, and hence continuous on S^2 . The definition of h_1 and φ_τ imply that $h_0 \circ F = \tilde{F} \circ h_1$ and $h_1(\tau) = \tilde{\tau}$ for each 1-cell $\tau \subset S^2$. We want to show next that h_1 is a homeomorphism onto \tilde{S}^2 .

Surjectivity is clear, because \tilde{S}^2 is equal to the union of all sets $\tilde{\tau} = h_1(\tau)$, where τ runs through all 1-cells.

To show injectivity, let $y \in \tilde{S}^2$ be arbitrary. There exists a 1-cell τ of minimal dimension such that $y \in \tilde{\tau}$. If σ is any other 1-cell with $y \in \tilde{\sigma}$, then $\tau \subset \sigma$. To see this, first note that by Lemma 13.10 we have $y \in \tilde{\tau} \cap \tilde{\sigma} = \widetilde{\tau \cap \sigma}$. Now if τ is not contained in σ , then it follows from Lemma 4.3 (i) that the set $\tau \cap \sigma$ is a union of 1-cells of dimension strictly less than the dimension of τ . The image of one of these 1-cells under the quotient map π has to contain y , contradicting the definition of τ .

Since $y \in \tilde{\tau} = h_1(\tau)$, there exists a point $x \in \tau$ with $h_1(x) = y$. We claim that y has no other preimage under h_1 . Indeed, suppose that $x' \in S^2$ and $h_1(x') = y$. Pick a 1-cell σ with $x' \in \sigma$. Then $y \in h_1(\sigma) = \tilde{\sigma}$, and so $\tau \subset \sigma$. Hence $x \in \sigma$. The map $h_1|_{\sigma} = \varphi_{\sigma}$ is injective on σ . Since $x, x' \in \sigma$ and $h_1(x) = y = h_1(x')$, we conclude $x = x'$ as desired. The injectivity of h_1 follows.

It remains to show that h_1 is orientation-preserving. Pick some positively-oriented flag (c_0, c_1, c_2) in \mathcal{D}^0 . Since \mathcal{D}^1 is a refinement of \mathcal{D}^0 , there exists a subflag in \mathcal{D}^1 , i.e., a flag (d_0, d_1, d_2) in \mathcal{D}^1 with $d_i \subset c_i$ for $i = 0, 1, 2$. Then (d_0, d_1, d_2) is also positively-oriented. The image of (c_0, c_1, c_2) under h_0 is the flag $(\tilde{c}_0, \tilde{c}_1, \tilde{c}_2)$. This flag is positively-oriented, because h_0 is orientation-preserving. The image of (d_0, d_1, d_2) under h_1 is the flag $(\tilde{d}_0, \tilde{d}_1, \tilde{d}_2)$. This is a subflag of $(\tilde{c}_0, \tilde{c}_1, \tilde{c}_2)$ and hence positively-oriented. It follows that h_1 maps one positively-oriented flag to a positively-oriented flag. Thus h_1 is orientation-preserving. \square

We are now ready to prove the main result of this section.

Proof of Proposition 13.1. Consider the map $\tilde{F}: \tilde{S}^2 \rightarrow \tilde{S}^2$ defined earlier. Since $\tilde{F} = h_0 \circ F \circ h_1^{-1}$, the map F is a branched covering map on S^2 , and h_1 and h_0 are orientation-preserving homeomorphisms, it follows that the map \tilde{F} is a branched covering map on \tilde{S}^2 . Moreover, we have $\text{crit}(\tilde{F}) = h_1(\text{crit}(F))$.

If v is a 1-vertex, then $\{v\}$ is a 1-cell of dimension 0, and so $h_1(\{v\}) = \widetilde{\{v\}} = \{\pi(v)\}$. Hence $h_1(v) = \pi(v)$. Since $F(\text{crit}(F)) \subset \text{post}(F)$, every point in $\text{crit}(\tilde{F})$ is a 1-vertex.

It follows that

$$\begin{aligned} \text{post}(\tilde{F}) &= \bigcup_{n \in \mathbb{N}} \tilde{F}^n(\text{crit}(\tilde{F})) = \bigcup_{n \in \mathbb{N}} \tilde{F}^n(h_1(\text{crit}(F))) \\ &= \bigcup_{n \in \mathbb{N}} \tilde{F}^n(\pi(\text{crit}(F))) = \bigcup_{n \in \mathbb{N}} (\pi \circ F^n)(\text{crit}(F)) = \pi(\text{post}(F)). \end{aligned}$$

In particular, $\text{post}(\tilde{F})$ is a finite set which implies that \tilde{F} is a Thurston map.

The Jordan curve $\tilde{\mathcal{C}} = \pi(\mathcal{C})$ satisfies

$$\text{post}(\tilde{F}) = \pi(\text{post}(F)) \subset \pi(\mathcal{C}) = \tilde{\mathcal{C}}.$$

Since $F(\mathcal{C}) \subset \mathcal{C}$, we also have

$$\tilde{F}(\tilde{\mathcal{C}}) = (\tilde{F} \circ \pi)(\mathcal{C}) = (\pi \circ F)(\mathcal{C}) \subset \pi(\mathcal{C}) = \tilde{\mathcal{C}}.$$

This shows that $\tilde{\mathcal{C}}$ is invariant with respect to \tilde{F} and contains the set of postcritical points of \tilde{F} .

The Jordan curve \mathcal{C} is the union of the 0-edges e . By definition of h_0 we have $h_0(e) = \tilde{e} = \pi(e)$ for each such 0-edge e and hence $h_0(\mathcal{C}) = \pi(\mathcal{C}) = \tilde{\mathcal{C}}$. Since \mathcal{D}^1 is a refinement of \mathcal{D}^0 , the curve \mathcal{C} can also be represented as a union of 1-edges. By Lemma 13.17 we have $h_1(e) = \tilde{e} = \pi(e)$ for each 1-edge e , and hence $h_1(\mathcal{C}) = \pi(\mathcal{C}) = \tilde{\mathcal{C}}$. Finally, every point in $\text{post}(F)$ is both a 0-vertex and a 1-vertex. This implies that $h_0(v) = \pi(v) = h_1(v)$ for all $v \in \text{post}(F)$. So h_0 and h_1 are orientation-preserving homeomorphism $S^2 \rightarrow \tilde{S}^2$ that map the Jordan curve \mathcal{C} to the same image $\tilde{\mathcal{C}}$ and agree on the finite set $\text{post}(F) \subset \mathcal{C}$. It follows that $h_1^{-1} \circ h_0$ is cellular for $(\mathcal{D}^0, \mathcal{D}^0)$. As in the proof of the second part of Lemma 12.3 this implies that $h_1^{-1} \circ h_0$ is isotopic to id_{S^2} rel. $\text{post}(F)$, and hence h_0 and h_1 are isotopic rel. $\text{post}(F)$. This implies that F and \tilde{F} are Thurston equivalent.

It remains to show that \tilde{F} is expanding. Since $\tilde{\mathcal{C}} \subset \tilde{S}^2$ is an \tilde{F} -invariant Jordan curve with $\text{post}(\tilde{F}) \subset \tilde{\mathcal{C}}$, we can do this by verifying the condition in Lemma 11.2.

Note that $\mathcal{D}^0(\tilde{F}, \tilde{\mathcal{C}}) = \tilde{\mathcal{D}}^0$. Moreover, since \tilde{F}^n is cellular for $(\tilde{\mathcal{D}}^n, \tilde{\mathcal{D}}^0)$, it follows from Lemma 5.4 that $\mathcal{D}^n(\tilde{F}, \tilde{\mathcal{C}}) = \tilde{\mathcal{D}}^n$ for all $n \in \mathbb{N}_0$. So the n -tiles for $(\tilde{F}, \tilde{\mathcal{C}})$ are precisely the sets \tilde{X} , where X is an n -tile on S^2 for (F, \mathcal{C}) . Moreover, for two tiles X and Y for (F, \mathcal{C}) we have $X \subset Y$ if and only if $\tilde{X} \subset \tilde{Y}$ (see Lemma 13.10). So if $Z^0 \supset Z^1 \supset Z^2 \dots$ is a nested sequences of n -tiles Z^n for $(\tilde{F}, \tilde{\mathcal{C}})$, then we can find a corresponding nested sequences $\{X^n\} \in \mathcal{S}$ such that $Z^n = \tilde{X}^n$ for all $n \in \mathbb{N}_0$. Thus, in order to establish that \tilde{F} is expanding, it suffices to show that if $\{X^n\} \in \mathcal{S}$ is arbitrary, then the intersection $\bigcap_n \tilde{X}^n$ contains precisely one point.

It is clear that this intersection contains at least one point. We argue by contradiction and assume that $\bigcap_n \tilde{X}^n$ contains more than one point, or equivalently, there exist two distinct (and hence disjoint) equivalence classes M and N with respect to \sim such that $M^n := M \cap X^n \neq \emptyset$ and $N^n := N \cap X^n \neq \emptyset$ for all $n \in \mathbb{N}_0$. Since equivalence classes and tiles are compact, in this way we get descending sequences $M^0 \supset M^1 \supset$

... and $N^0 \supset N^1 \supset \dots$ of nonempty and compact sets. Hence the sets $\bigcap_n M^n = M \cap \bigcap_n X^n$ and $\bigcap_n N^n = N \cap \bigcap_n X^n$ are nonempty. So there exists points $x \in M \cap \bigcap_n X^n$ and $y \in N \cap \bigcap_n X^n$. Since x and y lie in different equivalence classes, they are not equivalent. On the other hand, we have $x, y \in \bigcap_n X^n$ and $\{X^n\} \in \mathcal{S}$. Hence $x \sim y$ by Lemma 13.5 (ii). This is a contradiction showing that \tilde{F} is expanding. \square

Corollary 13.18. *Let $F: S^2 \rightarrow S^2$ be a Thurston map that has an invariant Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(F) \subset \mathcal{C}$. If F is combinatorially expanding for \mathcal{C} , then there exists an orientation-preserving homeomorphism $\phi: S^2 \rightarrow S^2$ that is isotopic to the identity on S^2 rel. $\text{post}(F)$ such that $\phi(\mathcal{C}) = \mathcal{C}$ and $G = \phi \circ F$ is an expanding Thurston map.*

Note that $\text{post}(G) = \text{post}(F)$ and $G(\mathcal{C}) = \mathcal{C}$. So the corollary says that if a Thurston map F is combinatorially expanding for an invariant Jordan curve $\mathcal{C} \subset \text{post}(F)$, then by “correcting” the map by post-composing with a suitable homeomorphism we can obtain an expanding Thurston map G with the same invariant curve and the same set of postcritical points.

Proof. Using the notation of Proposition 13.1 define $\phi = h_1^{-1} \circ h_0$. Then ϕ is isotopic to the identity on S^2 rel. $\text{post}(F)$ and $\phi(\mathcal{C}) = \mathcal{C}$. Moreover, $G = \phi \circ F = h_1^{-1} \circ \tilde{F} \circ h_1$, and so G is topologically conjugate to the expanding Thurston map \tilde{F} , and hence itself an expanding Thurston map. \square

Corollary 13.19. *Let $(\mathcal{D}^1, \mathcal{D}^0, L)$ be a two-tile subdivision rule on S^2 suppose that it can be realized by a Thurston map $F: S^2 \rightarrow S^2$ that is combinatorially expanding for the Jordan curve \mathcal{C} of \mathcal{D}^0 and for which its set $\text{post}(F)$ is equal to the set \mathbf{V}^0 of 0-vertices. Then the subdivision rule with the labeling can be realized by an expanding Thurston map G .*

In general we only have $\text{post}(F) \subset \mathbf{V}^0$. The stronger assumption $\text{post}(F) = \mathbf{V}^0$ is not very essential in the corollary. We added it for convenience in order to avoid some cumbersome technicalities in the proof.

Proof. Let G and ϕ be as in Corollary 13.18. Then $\text{post}(F) = \text{post}(G) = \mathbf{V}^0$, and so $\mathcal{D}^0 = \mathcal{D}^0(F, \mathcal{C}) = \mathcal{D}^0(G, \mathcal{C})$. Since ϕ is isotopic to the identity, this map is orientation-preserving. Since $\phi(\mathcal{C}) = \mathcal{C}$ and ϕ is the identity on $\text{post}(F) = \text{post}(G)$, the map ϕ is cellular for $(\mathcal{D}^0, \mathcal{D}^0)$. Since F is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$, the map $G = \phi \circ F$ is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$ and we have $G(c) = F(c)$ for each cell $c \in \mathcal{D}^1$. Since F realizes the

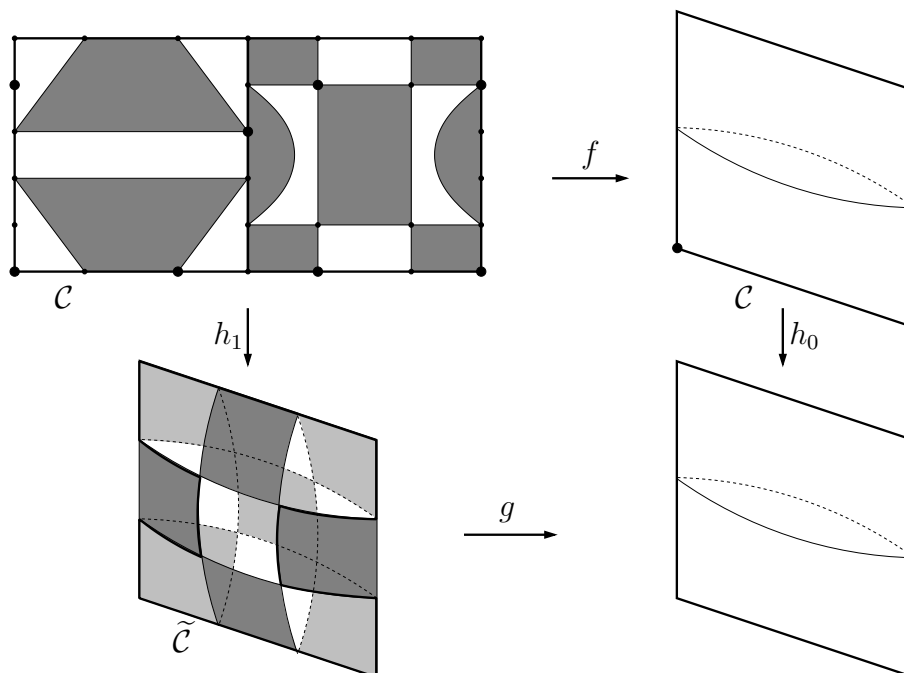


FIGURE 15. The map f is not combinatorially expanding, but equivalent to the expanding map g .

subdivision rule with the given labeling, this shows that G is also a realization. Since G is expanding, the claim follows. \square

Proposition 13.1 shows that for a Thurston map with an invariant Jordan curve containing all its postcritical points, combinatorial expansion is *sufficient* for the existence of a Thurston equivalent map that is expanding. The question arises whether combinatorial expansion is *necessary* as well. The answer is negative, as the following example of a Thurston map f will show. The map f has an invariant Jordan curve \mathcal{C} containing all its postcritical points. It is not combinatorially expanding for \mathcal{C} (and hence not expanding), yet equivalent to an expanding Thurston map g .

Example 13.20. The map $f: S^2 \rightarrow S^2$ is illustrated in the top of Figure 15. It has four postcritical points, which are the vertices of the pillow shown in the top right. The front shows the white 0-tile, the back is the black 0-tile. The subdivision of the 0-tiles is indicated in the top left of the figure. Here we have cut the pillow along three 0-edges. The left shows the subdivision of the white 0-tile, the right the subdivision of the black 0-tile. On the left the preimages of one postcritical point (a 0-vertex) are given. So we only show the labels of

1-vertices that the labeling sends to the 0-vertex on the bottom left of the pillow.

The boundary of the pillow is the Jordan curve \mathcal{C} containing all postcritical points of f . It is invariant under f , but the map f is not combinatorially expanding for \mathcal{C} : for arbitrarily large n there will always be an n -tile (contained in the white 0-tile) joining the 0-edges on the left and on the right.

We will see that f is Thurston equivalent to the map $g: \tilde{S}^2 \rightarrow \tilde{S}^2$, indicated in the bottom of Figure 15. The map g is a Lattès map, obtained similarly as the example in Section 1.2 as a quotient of $\psi: \mathbb{C} \rightarrow \mathbb{C}$, $\psi(z) = 3z$. It is clear that g is expanding.

We fix a homeomorphism $h_0: S^2 \rightarrow \tilde{S}^2$ that maps the boundary of the pillow of f (i.e., the curve \mathcal{C}) to the boundary of the pillow of g , and the postcritical points of f to the postcritical points of g . Let \mathcal{D}^1 and $\tilde{\mathcal{D}}^1$ denote the cell decompositions of S^2 and \tilde{S}^2 , respectively, into 1-cells as indicated on the left of Figure 15. As in the proof of Proposition 12.3 there is a homeomorphism $h_1: S^2 \rightarrow \tilde{S}^2$ that maps the curve $\mathcal{C} \subset S^2$ to the curve $\tilde{\mathcal{C}} \subset \tilde{S}^2$ indicated in the figure, is cellular with respect to $(\mathcal{D}^1, \tilde{\mathcal{D}}^1)$, and satisfies

$$g \circ h_1 = h_0 \circ f.$$

Obviously, the map h_1 is isotopic to h_0 rel. $\text{post}(f)$. Thus f and g are Thurston equivalent.

Note that g has an invariant curve $\tilde{\mathcal{C}}'$ (different from $\tilde{\mathcal{C}}$ of course) that contains all its postcritical points.

14. AUXILIARY RESULTS ON GRAPHS

The main result in this section is Lemma 14.5 which gives a criterion when a Jordan curve \mathcal{C} in a 2-sphere S^2 can be isotoped relative to a finite set $P \subset \mathcal{C}$ into the 1-skeleton of a given cell decomposition \mathcal{D} of S^2 . We first discuss some facts about graphs that are needed in the proof. Since all the graphs we consider will be embedded in a 2-sphere, we base the concept of a graph on a topological definition rather than a combinatorial one as usual.

A (*finite*) *graph* is a compact Hausdorff space G equipped with a fixed cell decomposition \mathcal{D} such that $\dim(c) \leq 1$ for all $c \in \mathcal{D}$. The cells c in \mathcal{D} of dimension 1 are called the *edges* of the graph, and the points $v \in G$ such that $\{v\}$ is a 0-dimensional cell in \mathcal{D} the *vertices* of the graph.

An *oriented edge* e in a graph is an edge, where one of the vertices in ∂e has been chosen as the *initial point* and the other vertex as the

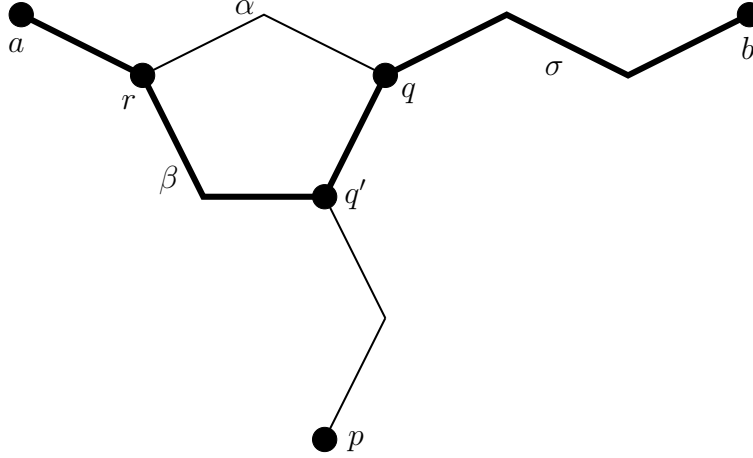
terminal point of e . An *edge path* in G is a finite sequence α of oriented edges e_1, \dots, e_N such that the terminal point of e_i is the initial point of e_{i+1} for $i = 1, \dots, N - 1$. We denote by $|\alpha| = e_1 \cup \dots \cup e_N$ the underlying set of the edge path. The edge path α *joins* the vertices $a, b \in G$ if the initial point of e_1 is a and the terminal point of e_N is b . The number N is called the *length* of the edge path. The edge path is called *simple* if e_i and e_j are disjoint for $1 \leq i < j \leq N$ and $j - i \geq 2$, and $e_i \cap e_j$ consists of precisely one point (the terminal point of e_i and initial point of e_j) when $j = i + 1$. If the edge path α is simple, then $|\alpha|$ is an arc. The edge path is called a *loop* if the terminal point of e_N is the initial point of e_1 .

A graph is connected (as a topological space) if and only if any two vertices $a, b \in G$, $a \neq b$, can be joined by an edge path. The *combinatorial distance* of two vertices a and b in a connected graph G is defined as the minimal length of all edge paths joining the points (interpreted as 0 if $a = b$). The vertices $a, b \in G$ are called *neighbors* if their combinatorial distance is equal to 1, i.e., if there exists an edge e in G whose endpoints are a and b . A vertex $q \in G$ is called a *cut point* of G if $G \setminus \{q\}$ is not connected. A vertex $q \in G$ is not a cut point if and only if all vertices $a, b \in G \setminus \{q\}$, $a \neq b$, can be joined by an edge path α with $q \notin |\alpha|$.

Lemma 14.1. *Let G be a connected graph without cut points. Then for all vertices $a, b, p \in G$ with $a \neq b$ there exists a simple edge path γ in G with $p \in |\gamma|$ that joins a and b .*

Proof. Since G is connected, there exist edge paths in G joining a and b . By removing loops from such a path if necessary, we can also obtain such an edge path in G that is simple. Among all such simple paths, there is one that contains a vertex with minimal combinatorial distance to p . More precisely, there exists a simple edge path α in G with endpoints a and b , and a vertex $q \in |\alpha|$ such that the combinatorial distance $k \in \mathbb{N}_0$ of q and p is minimal among all combinatorial distances between p and vertices on simple paths joining a and b . If $k = 0$ then $q = p$ and we can take $\gamma = \alpha$.

We will show that the alternative case $k \geq 1$ leads to a contradiction. By definition of combinatorial distance, there exists an edge path joining q to p consisting of $k \geq 1$ edges. The second vertex q' on this path as travelling from q to p is a neighbor of q whose combinatorial distance to p is $k - 1$ and hence strictly smaller than the combinatorial distance of q to p . In particular, $q' \notin |\alpha|$ by choice of q and α . We will obtain the desired contradiction if we can show that there exists a simple edge path σ in G that joins a and b and passes through q' .

FIGURE 16. Constructing a path through a, b, p .

For the construction of σ we apply our assumption that G has no cut points; in particular, q is no cut point and hence there exists an edge path β with $q \notin |\beta|$ that joins q' to a vertex in the (nonempty) set $A = |\alpha| \setminus \{q\}$. We may assume that β is simple and that the endpoint $r \neq q'$ of β is the only point in $|\beta| \cap A$.

Moreover, we may assume that r lies between a and q on the path α (the argument in the other case where r lies between q and b is similar). Now let σ be the edge path obtained by traveling from a to r along α , then from r to q' along β , then from q' to q along an edge (this is possible since q and q' are neighbors), and finally from q to b along α . See the illustration in Figure 16. Then σ is a simple edge path in G that passes through q' and has the endpoints a and b . This gives the desired contradiction. \square

Now let S^2 be a 2-sphere, and \mathcal{D} be a cell decomposition of S^2 . We denote the set of tiles, edges, and vertices in \mathcal{D} by \mathbf{X} , \mathbf{E} , and \mathbf{V} , respectively. In the following the term cell, tile, etc., refers to elements in these sets.

Let $M \subset \mathbf{X}$ be a set of tiles. We denote by $|M|$ its underlying set; so

$$|M| = \bigcup_{X \in M} X.$$

The set

$$(14.1) \quad G_M := \bigcup_{X \in M} \partial X$$

admits a natural cell decomposition consisting of all cells contained in G_M . Obviously, no such cell can be a tile, so with this cell decomposition G_M is a graph.

Similarly as in Section 12, we call a sequence of tiles $X = X_1, \dots, X_N = Y$ an *e-chain* joining X and Y if for $i = 1, \dots, N$ there exists an edge e_i with $e_i \subset \partial X_{i-1} \cap \partial X_i$. A set M of tiles is *e-connected* if every two tiles in M can be joined by an *e-chain*.

Lemma 14.2. *Let $M \subset \mathbf{X}$ be a set of tiles that is e-connected. Then the graph G_M is connected and has no cut points.*

Proof. Let $a, b \in G_M$ be arbitrary vertices with $a \neq b$. We can pick tiles X and Y in M such that a is a vertex in X and b is a vertex in Y . By assumption there exists an *e-chain* X_1, X_1, \dots, X_N in M with $X_1 = X$ and $X_N = Y$. The vertices of a tile X_i lie in G_M ; they subdivide the Jordan curve ∂X_i such that successive vertices on ∂X_i are connected by an edge and are hence neighbors in G_M . An edge path α in G_M joining a and b can now be obtained as follows: starting from $a \in \partial X_0$, use edges on the boundary of X_0 to find an edge path in G_M that joins $p_0 = a$ to a vertex p_1 of X_1 . This is possible, since X_0 and X_1 have a common edge and hence at least two common vertices. Then run from p_1 along edges on ∂X_1 to a vertex p_2 of X_2 , and so on. Once we arrived at a vertex p_k of X_k , we can reach b by running from p_k to $p_{k+1} = b$ along edges on X_k . In this way we obtain an edge path α in G_M that joins a and b .

A slight refinement of this argument also shows that we can construct the path α so that it avoids any given vertex q in G_M distinct from a and b . Indeed, choose $p_0 = a$ as before. Since X_0 and X_1 have at least two vertices in common, we can pick a common vertex p_1 of X_0 and X_1 that is distinct from q . There exists an arc on ∂X_0 (possibly degenerate) that does not contain q and joins p_0 and p_1 . This arc (if non-degenerate) consists of edges and if we follow these edges, we obtain an edge path in G_M that does not contain q and joins p_0 and p_1 . In the same way we can find an edge path in G_M that avoids q and joins p_1 to a vertex $p_2 \in \partial X_1 \cap \partial X_2$, and so on. Concatenating all these edge paths we get a path α as desired.

This shows that G_M is connected and has no cut points. \square

Lemma 14.3. *Let $M \subset \mathbf{X}$ be a set of tiles that is e-connected, and let $a, b, p \in |M|$ be distinct vertices. Then there exists a simple edge path α in G_M with $p \in |\alpha|$ that joins a and b .*

In particular, this applies if M consists of a single *e-chain*.

Proof. This follows from Lemma 14.2 and Lemma 14.1. \square

Lemma 14.4. *Let $\gamma: J \rightarrow S^2$ be a path in S^2 defined on a closed interval $J \subset \mathbb{R}$ and $M = M(\gamma)$ be the set of tiles having nonempty intersection with γ . Then M is e -connected.*

Proof. We first prove the following claim. If $[a, b] \subset \mathbb{R}$, $\alpha: [a, b] \rightarrow S^2$ is a path, and X and Y are tiles with $\alpha(a) \in X$ and $\alpha(b) \in Y$, then there exists an e -chain $X_1 = X, X_2, \dots, X_N = Y$ such that $X_i \cap \alpha \neq \emptyset$ for all $i = 1, \dots, N$.

In the proof of this claim, we call an e -chain X_1, \dots, X_N *admissible* if $X_1 = X$ and $X_i \cap \alpha \neq \emptyset$ for all $i = 1, \dots, N$. So we want to find an admissible e -chain whose last tile is Y .

Let $T \subset [a, b]$ be the set of all points $t \in [a, b]$ for which there exists an admissible e -chain X_1, \dots, X_N with $\alpha(t) \in X_N$. We first want to show that $b \in T$.

First note that the set T is closed. Indeed, suppose that (t_k) is a sequence in T with $t_k \rightarrow t_\infty \in [a, b]$ as $k \rightarrow \infty$. Then for each $k \in \mathbb{N}$ there exists an admissible e -chain $X_1^k, \dots, X_{N_k}^k$ with $\alpha(t_k) \in X_{N_k}^k$. Define $Z_k = X_{N_k}^k$ to be the last tile in this chain. Since there are only finitely many tiles, there exists one tile, say Z , among the tiles Z_1, Z_2, Z_3, \dots that appears infinitely often in this sequence. Since $\alpha(t_k) \rightarrow \alpha(t_\infty)$, we have $\alpha(t_k) \in Z$ for infinitely many k , and tiles are closed, we conclude $\alpha(t_\infty) \in Z$. By definition of Z there exists an admissible e -chain X_1, \dots, X_N with $X_N = Z$. Then $\alpha(t_\infty) \in Z = X_N$, and so $t_\infty \in T$.

Obviously, $a \in T$ and so T is nonempty. Since T is also closed, the set T has a maximum, say $m \in [a, b]$. We have to show that $m = b$; we will see that the assumption $m < b$ leads to a contradiction.

Consider $p = \alpha(m)$. Then there exists an admissible e -chain X_1, \dots, X_N with $p \in Z := X_N$.

If $p \in \text{int}(Z)$, then $\alpha(t) \in Z$ and so $t \in T$ for $t \in (m, b]$ close to m . This is impossible by definition of m .

If p does not belong to $\text{int}(Z)$, then p must be a boundary point of Z . Suppose first that p is in the interior of an edge $e \subset \partial Z$. By Lemma 5.1 (iv) there exists precisely one tile Z' distinct from Z such that $e \subset \partial Z'$. Moreover, $Z \cup Z'$ is a neighborhood of p , and so points $\alpha(t)$ with $t \in (m, b]$ close to m belong to Z or Z' . Since Z' contains p and hence meets α , and Z and Z' share an edge, X_1, \dots, X_N, Z' is an admissible e -chain. It follows that $t \in T$ for $t \in (m, b]$ close to m . Again this is impossible by definition of m .

If p is a boundary point of Z , but not in the interior of an edge, then p is a vertex. The tiles in the cycle of p form a neighborhood of p , and so a point $\alpha(t)$ for some $t \in (m, b]$ close to m will belong to a tile Z'

in the cycle of p . It follows from Lemma 5.1 that any two tiles in the cycle of a vertex can be connected by an e -chain consisting of tiles in the cycle. Hence there exists an e -chain $Z = Z_1, \dots, Z_K = Z'$ such that $p \in Z_j$ for $j = 1, \dots, K$. In particular, $\alpha \cap Z_j \neq \emptyset$ for $j = 1, \dots, K$, and so $X_1, \dots, X_N = Z = Z_1, \dots, Z_K = Y'$ is an admissible e -chain. Since $\alpha(t) \in Z' = Z_K$, we have $t \in T$, again a contradiction.

We have exhausted all possibilities proving that $b \in T$ as desired. So there exists an admissible e -chain $X_1 = X, \dots, X_N$ with $\alpha(b) \in X_N$. If $X_N = Y$, then we are done. If $X_N \neq Y$, then $\alpha(b) \in \partial X_N \cap \partial Y$, and so $\alpha(b)$ is an interior point of an edge e with $e \subset \partial X_N \cap \partial Y$, or $\alpha(b)$ is a vertex. As in the first part of the proof, one then can extend the admissible e -chain $X_1 = X, \dots, X_N$ to obtain an admissible e -chain whose last tile is Y . The claim follows.

The claim now easily implies the statement of the lemma. Indeed, pick a point $a \in J$ and fix a tile $X \in \mathbf{X}$ with $\gamma(a) \in X$. Then $X \in M = M(\gamma)$. If $Y \in M$ is arbitrary, then there exists $b \in J$ with $\gamma(b) \in Y$. If $b \geq a$, then we apply the claim to the path $\alpha = \gamma| [a, b]$, and if $b \leq a$ to the path $\alpha = \gamma| [b, a]$. This shows that we can find an e -chain in M that joins $Y \in M$ to the fixed tile X . Hence any two tiles in M can be joined by an e -chain in M . \square

For the formulation of the next statement, we need a slight generalization of Definition 7.6. Let $\mathcal{C} \subset S^2$ be a Jordan curve, and $P \subset \mathcal{C}$ be a finite set with $\#P \geq 3$. The points in P divide \mathcal{C} into subarcs that have endpoints in P , but whose interiors are disjoint from P . We say that a set $K \subset S^2$ *joins opposite sides of* (\mathcal{C}, P) if $k = \# \text{post}(f) \geq 4$ and K meets two of these arcs that are non-adjacent (i.e., disjoint), or if $k = \# \text{post}(f) = 3$ and K meets all of these arcs (in this case there are three arcs).

In the following proposition and its proof, metric notions refer to some fixed base metric on S^2 .

Lemma 14.5. *Let $\mathcal{C} \subset S^2$ be a Jordan curve, and $P \subset \mathcal{C}$ a finite set with $k = \#P \geq 3$. Then there exists $\epsilon_0 > 0$ with the following property:*

Suppose that \mathcal{D} is a cell decomposition of S^2 with vertex set \mathbf{V} and 1-skeleton E . If $P \subset \mathbf{V}$ and

$$\max_{c \in \mathcal{D}} \text{diam}(c) < \epsilon_0,$$

then there exists a Jordan curve $\mathcal{C}' \subset E$ that is isotopic to \mathcal{C} rel. P and so that no tile in \mathcal{D} joins opposite sides of (\mathcal{C}', P) .

Proof. Fix an orientation of \mathcal{C} and let p_1, \dots, p_k be the points in P in cyclic order on \mathcal{C} . The points in P divide \mathcal{C} into subarcs $\mathcal{C}_1, \dots, \mathcal{C}_k$

such that for $i = 1, \dots, k$ the arc \mathcal{C}_i has the endpoints p_i and p_{i+1} and has interior disjoint from P . Here and in the following the index i is understood modulo k , i.e., $p_{k+1} = p_1$, etc. Note that $\mathcal{C}_i \cap \mathcal{C}_{i+1} = \{p_{i+1}\}$ for $i = 1, \dots, k$. There exists a number $\delta_0 > 0$ such that no set $K \subset S^2$ with $\text{diam}(K) < \delta_0$ joins opposite sides of (\mathcal{C}, P) (this can be seen as in the discussion after (7.3)).

Now choose $\delta > 0$ as in Proposition 10.7 for $J = \mathcal{C}$ and $n = k$. We may assume that $3\delta < \delta_0$.

We break up \mathcal{C} into subarcs

$$(14.2) \quad \alpha_1, \gamma_1, \alpha_2, \gamma_2, \dots, \gamma_k, \alpha_1,$$

arranged in cyclic order on \mathcal{C} , such that p_i is an interior point of α_i and we have $\alpha_i \subset B(p_i, \delta/2)$ for each $i = 1, \dots, k$. The arcs in (14.2) have disjoint interiors, and two arcs have an endpoint in common if and only if there are adjacent in this cyclic order in which case they share one endpoint. So each “middle piece” γ_i does not contain any point from P and is contained in the interior of \mathcal{C}_i .

We choose $0 < \epsilon_0 < \delta/4$ so small that the distance between non-adjacent arcs in (14.2) is $\geq 10\epsilon_0$ and so that

$$\text{dist}(p_i, \gamma_{i-1} \cup \gamma_i) \geq 10\epsilon_0$$

for $i = 1, \dots, k$.

Now suppose we have a cell decomposition \mathcal{D} of S^2 such that P is contained in the vertex set \mathbf{V} of \mathcal{D} and

$$\max_{c \in \mathcal{D}} \text{diam}(c) < \epsilon_0.$$

Our goal is to find a Jordan curve $\mathcal{C}' \subset S^2$ consisting of arcs \mathcal{C}'_i that are unions of edges, have endpoints p_i and p_{i+1} , and satisfy

$$\mathcal{C}'_i \subset \mathcal{N}^\delta(\mathcal{C}_i)$$

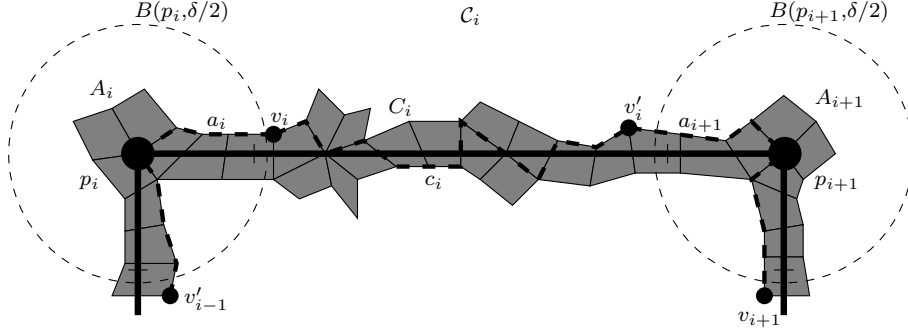
for $i = 1, \dots, k$.

Let \mathbf{A}_i be the set of all tiles intersecting α_i and \mathbf{C}_i be the set of all tiles intersecting γ_i , for all $i = 1, \dots, k$. Recall that for a given set of tiles M , we denote by $|M|$ the union of tiles in M . Let $A_i := |\mathbf{A}_i|$ and $C_i = |\mathbf{C}_i|$. Note that

$$A_i \subset \mathcal{N}^{\epsilon_0}(\alpha_i) \quad \text{and} \quad C_i \subset \mathcal{N}^{\epsilon_0}(\gamma_i).$$

Furthermore

$$(14.3) \quad \begin{aligned} A_i \cup C_i \cup A_{i+1} &\subset \mathcal{N}^{\epsilon_0}(\alpha_i) \cup \mathcal{N}^{\epsilon_0}(\gamma_i) \cup \mathcal{N}^{\epsilon_0}(\alpha_{i+1}) \\ &\subset B(p_i, \delta) \cup \mathcal{N}^{\epsilon_0}(\gamma_i) \cup B(p_{i+1}, \delta) \\ &\subset \mathcal{N}^\delta(\mathcal{C}_i), \end{aligned}$$

FIGURE 17. Construction of the curve \mathcal{C}' .

and the natural cyclic order of these sets is

$$(14.4) \quad A_1, C_1, A_2, C_2, \dots, A_k, C_k, A_1.$$

By choice of ϵ_0 we have that if two of the sets in (14.4) are not adjacent in the cyclic order, then their distance is $\geq 8\epsilon_0$ and so their intersection is empty. Moreover, for $i = 1, \dots, k$ the only one of these sets that contains p_i is A_i . The construction that follows is illustrated in Figure 17. Here the two large dots represent two points p_i, p_{i+1} and the thick line the curve \mathcal{C} .

Consider the graphs G_{A_i}, G_{C_i} of all edges in A_i, C_i , see (14.1). Note that there is at least one tile contained in both A_i and C_i , namely any tile containing the point where α_i, γ_i intersect. Thus G_{A_i} and G_{C_i} , similarly G_{C_i} and $G_{A_{i+1}}$, are not disjoint.

Pick a simple edge path c'_i in G_{C_i} connecting A_i, A_{i+1} . Let c_i be a subarc of $|c'_i|$ whose initial point v_i lies in A_i , whose terminal point v'_i lies in A_{i+1} , and that has no other points in common with A_i and A_{i+1} . The points v_i and v'_i are vertices. To see this, suppose that v_i , for example, is not a vertex. Then, since $v_i \in c'_i$ and c'_i is an edge path, there exists an edge $e \subset c_i$ such that $v_i \in \text{int}(e)$. But then necessarily $e \subset A_i$ and v_i cannot be the only point of c_i that belongs to A_i .

Note that $v'_{i-1} \in C_{i-1}$ and $v_i \in C_i$ are distinct vertices in A_i . Recall that $p_i \in A_i$. Thus it follows from Lemma 14.4 and Lemma 14.3 that there exists an arc $a_i \subset A_i$ with $p_i \in a_i$ that consists of edges and has the endpoints v'_{i-1} and v_i . Since $p_i \notin C_{i-1} \cup C_i$, we have $v'_{i-1}, v_i \neq p_i$, and so $p_i \in \text{int}(a_i)$.

If we arrange the arcs a_i and c_i in cyclic order

$$a_1, c_1, a_2, c_2, \dots, a_k, c_k, a_1,$$

then two of these arcs have nonempty intersection if and only if they are adjacent in the order. If two arcs are adjacent, then their intersection

consists of a common endpoint. Therefore, the set

$$\mathcal{C}' := a_1 \cup c_1 \cup a_2 \cup c_2 \cup \cdots \cup a_k \cup c_k$$

is a Jordan curve that passes through the points p_1, \dots, p_k . Moreover, \mathcal{C}' consists of edges and is hence contained in the 1-skeleton E of \mathcal{D} .

By construction each vertex p_i is an interior point of the arc a_i . Thus it divides a_i into two subarcs a_i^- and a_i^+ consisting of edges such that p_i is a common endpoint of a_i^- and a_i^+ , and such that a_i^- shares an endpoint with c_{i-1} and a_i^+ one with c_i . Then

$$\mathcal{C}'_i := a_i^+ \cup c_i \cup a_{i+1}^-$$

for $i = 1, \dots, k$ is an arc that consists of edges and has endpoints p_i and p_{i+1} . The arcs $\mathcal{C}'_1, \dots, \mathcal{C}'_k$ have pairwise disjoint interior. Moreover,

$$\mathcal{C}' = \mathcal{C}'_1 \cup \cdots \cup \mathcal{C}'_k,$$

and by (14.3) we have

$$\mathcal{C}'_i \subset A_i \cup C_i \cup A_{i+1} \subset \mathcal{N}^\delta(\mathcal{C}_i).$$

Hence by Proposition 10.7 and choice of δ , the curve \mathcal{C}' is isotopic to \mathcal{C} rel. P .

It remains to show that no tile in \mathcal{D} joins opposite sides of (\mathcal{C}', P) . To see this, we argue by contradiction. Suppose that there exists a tile X in \mathcal{D} that joins opposite sides of (\mathcal{C}', P) . Then $K := \mathcal{N}^\delta(X)$ joins opposite sides of (\mathcal{C}, P) , since $\mathcal{C}'_i \subset \mathcal{N}^\delta(\mathcal{C}_i)$ for all $i = 1, \dots, k$. By choice of δ_0 we then have

$$\delta_0 \leq \text{diam}(K) \leq 2\delta + \text{diam}(X) \leq 2\delta + \epsilon_0 < 3\delta < \delta_0,$$

which is impossible. \square

15. INVARIANT CURVES

This section is central for this work. We will prove existence and uniqueness results for invariant Jordan curves of expanding Thurston maps. We will also show that if an invariant Jordan curve exists, then it can be obtained from an iterative procedure. We start by looking at a specific example that will illustrate some of the main ideas.

Example 15.1. Let $S^2 = \widehat{\mathbb{C}}$ and $f: S^2 \rightarrow S^2$ be the map defined by $f(z) = 1 + (\omega - 1)/z^3$ for $z \in S^2$, where $\omega = e^{4\pi i/3}$. Note that $f = f_6$ is one of the maps considered in Example 12.14. It realizes the subdivision rule shown in Figure 12 and Figure 13.

Note that $f(z) = \tau(z^3)$, where $\tau(w) = 1 + (\omega - 1)/w$ is a Möbius transformation that maps the upper half-plane to the half-plane above

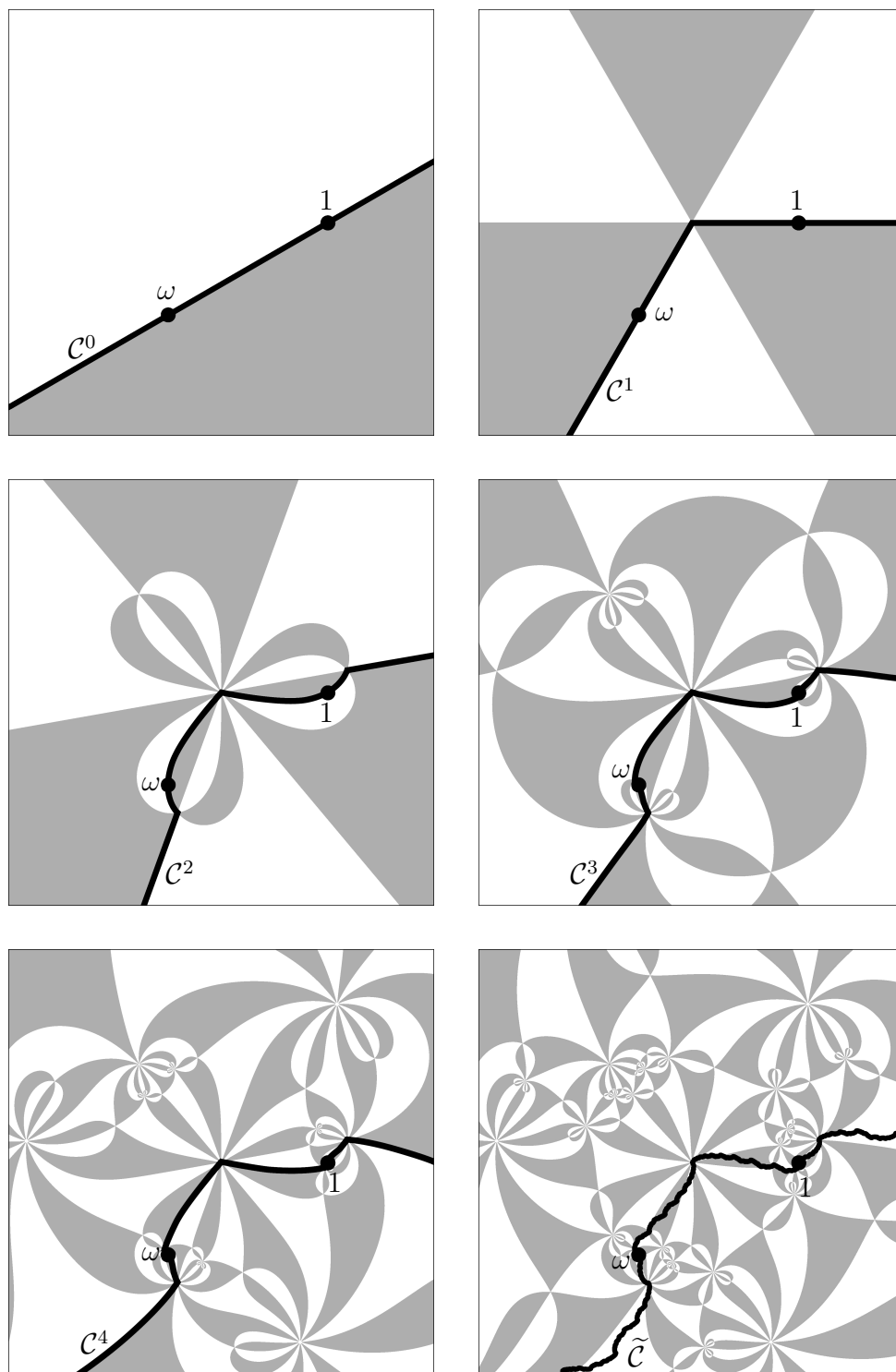


FIGURE 18. The invariant curve for Example 15.1.

the line through the points ω and 1 (indeed, τ maps $0, 1, \infty$ to $\infty, \omega, 1$, respectively). We have $\text{crit}(f) = \{0, \infty\}$ and $\text{post}(f) = \{\omega, 1, \infty\}$.

One can obtain an f -invariant Jordan curve $\tilde{\mathcal{C}} \subset S^2$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$ as follows. We start by choosing a Jordan curve $\mathcal{C}^0 \subset S^2$ containing all postcritical points of f . Here we let \mathcal{C}^0 be the (extended) line through ω and 1 (i.e., the circle on $\hat{\mathbb{C}}$ through $\omega, 1, \infty$).

Now consider $f^{-1}(\mathcal{C}^0) = \bigcup_{k=0, \dots, 5} R_k$, where

$$R_k = \{re^{ik\pi/3} : 0 \leq r \leq \infty\}$$

is the ray from 0 through the sixth root of unity $e^{ik\pi/3}$; see the top right in Figure 18. We choose a Jordan curve $\mathcal{C}^1 \subset S^2$ such that

$$\mathcal{C}^1 \subset f^{-1}(\mathcal{C}^0), \text{ post}(f) \subset \mathcal{C}^1, \text{ and } \mathcal{C}^1 \text{ is isotopic to } \mathcal{C}^0 \text{ rel. post}(f).$$

This is not always possible, but in our specific case there is a unique Jordan curve $\mathcal{C}^1 \subset f^{-1}(\mathcal{C}^0)$ with $\text{post}(f) \subset \mathcal{C}^1$, namely $\mathcal{C}^1 = R_0 \cup R_4$, the union of the two rays through ω and through 1. Since $\#\text{post}(f) = 3$, the requirement that \mathcal{C}^1 is isotopic to \mathcal{C}^0 rel. $\text{post}(f)$ is automatic for our specific map f by Lemma 10.10. Let $H: S^2 \times I \rightarrow S^2$ be an isotopy rel. $\text{post}(f)$ that deforms \mathcal{C}^0 to \mathcal{C}^1 , i.e., $H_0 = \text{id}_{S^2}$ and $H_1(\mathcal{C}^0) = \mathcal{C}^1$.

Given the data \mathcal{C}^0 , \mathcal{C}^1 , and H , there are two ways to obtain an f -invariant Jordan curve isotopic to \mathcal{C}^0 rel. $\text{post}(f)$ and isotopic to \mathcal{C}^1 rel. $f^{-1}(\text{post}(f))$.

For the first approach we consider the Thurston map $\hat{f} := H_1 \circ f$. Since $\mathcal{C}^1 \subset f^{-1}(\mathcal{C}^0)$ we have $f(\mathcal{C}^1) \subset \mathcal{C}^0$, and so

$$\hat{f}(\mathcal{C}^1) = (H_1 \circ f)(\mathcal{C}^1) \subset H_1(\mathcal{C}^0) = \mathcal{C}^1.$$

Thus \mathcal{C}^1 is \hat{f} -invariant. The two-tile subdivision rule given by $\mathcal{D}^1 = \mathcal{D}^1(\hat{f}, \mathcal{C}^1)$, $\mathcal{D}^0 = \mathcal{D}^0(\hat{f}, \mathcal{C}^1)$, and the labeling induced by \hat{f} is as in Figure 13. The map \hat{f} is combinatorially expanding for \mathcal{C}^1 ; indeed, no 2-tile for (\hat{f}, \mathcal{C}^1) joins opposite sides of \mathcal{C}^1 . Thus by Corollary 13.18 there is a homeomorphism $\phi: S^2 \rightarrow S^2$ isotopic to the identity on S^2 rel. $\text{post}(\hat{f}) = \text{post}(f)$ such that $\phi(\mathcal{C}^1) = \mathcal{C}^1$ and $g = \phi \circ \hat{f}$ is expanding. Since f is also expanding (this follows from Proposition 19.1 below) and g is Thurston equivalent to f , there is a homeomorphism $h: S^2 \rightarrow S^2$ such that $h \circ f = g \circ h$ (Theorem 10.4). Then $\tilde{\mathcal{C}} := h^{-1}(\mathcal{C}^1)$ is an f -invariant Jordan curve containing $\text{post}(f)$. In Theorem 1.3 we give a general existence result, which is proved in the same fashion.

For the second approach we use Proposition 10.1 to lift $H = H^0$ by the map f to an isotopy H^1 with $H_0^1 = \text{id}_{S^2}$. Then we lift H^1 to an isotopy H^2 with $H_0^2 = \text{id}_{S^2}$, etc. In this way, we find a sequence of isotopies H^n and inductively define $\mathcal{C}^{n+1} := H_1^n(\mathcal{C}^n)$. We will see in

Proposition 15.14 that the sequence $\{\mathcal{C}^n\}$ of Jordan curves converges in the Hausdorff sense to an f -invariant Jordan curve $\tilde{\mathcal{C}}$ containing all postcritical points of f as desired. This is illustrated in Figure 18.

In fact in our example there is a *unique* f -invariant Jordan curve $\tilde{\mathcal{C}} \subset \hat{\mathcal{C}}$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$. To see this, note that since $\# \text{post}(f) = 3$, every such curve $\tilde{\mathcal{C}}$ is isotopic rel. $\text{post}(f)$ to the curve \mathcal{C}^0 chosen above. So we can find an isotopy $K: S^2 \times I \rightarrow S^2$ rel. $\text{post}(f)$ with $K_0 = \text{id}_{S^2}$ and $K_1(\tilde{\mathcal{C}}) = \mathcal{C}^0$. By Proposition 10.1 we can lift K to an isotopy $\tilde{K}: S^2 \times I \rightarrow S^2$ rel. $f^{-1}(\text{post}(f))$ with $\tilde{K}_0 = \text{id}_{S^2}$ and $K_t \circ f = f \circ \tilde{K}_t$ for $t \in I$. Then by Lemma 10.2 we have

$$\mathcal{C}' := \tilde{K}_1(\tilde{\mathcal{C}}) \subset \tilde{K}_1(f^{-1}(\tilde{\mathcal{C}})) = f^{-1}(K_1(\tilde{\mathcal{C}})) = f^{-1}(\mathcal{C}^0).$$

So \mathcal{C}' is a Jordan curve in S^2 with $\mathcal{C}' \subset f^{-1}(\mathcal{C}^0)$ and $\text{post}(f) \subset \mathcal{C}'$. Since \mathcal{C}^1 is the unique such curve, we conclude $\mathcal{C}' = \tilde{K}_1(\tilde{\mathcal{C}}) = \mathcal{C}^1$. In particular, $\tilde{\mathcal{C}}$ is isotopic to \mathcal{C}^1 rel. $f^{-1}(\text{post}(f))$ by the isotopy \tilde{K} . So every f -invariant Jordan curve $\tilde{\mathcal{C}}$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$ lies in the same isotopy class rel. $f^{-1}(\text{post}(f))$ as \mathcal{C}^1 . Hence by Theorem 1.4 (which we will prove momentarily) there is at most one such Jordan curve $\tilde{\mathcal{C}}$. The uniqueness of $\tilde{\mathcal{C}}$ follows.

15.1. Existence and uniqueness of invariant curves. We start with establishing uniqueness results.

Proof of Theorem 1.4. Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map, and suppose that \mathcal{C} and \mathcal{C}' are f -invariant Jordan curves in S^2 that both contain the set $\text{post}(f)$ and are isotopic rel. $f^{-1}(\text{post}(f))$. We have to show that $\mathcal{C} = \mathcal{C}'$.

Under the given assumptions there exists an isotopy $H^0: S^2 \times I \rightarrow S^2$ rel. $f^{-1}(\text{post}(f))$ with $H_0^0 = \text{id}_{S^2}$ and $H_1^0(\mathcal{C}) = \mathcal{C}'$. Since $\text{post}(f) \subset f^{-1}(\text{post}(f))$, the map H^0 is also an isotopy rel. $\text{post}(f)$. Hence by Proposition 10.1 we can find an isotopy $H^1: S^2 \times I \rightarrow S^2$ rel. $f^{-1}(\text{post}(f))$ with $H_0^1 = \text{id}_{S^2}$ and $f \circ H_t^1 = H_t^0 \circ f$ for all $t \in I$. Repeating this argument, we obtain isotopies $H^n: S^2 \times I \rightarrow S^2$ rel. $f^{-1}(\text{post}(f))$ with $H_0^n = \text{id}_{S^2}$ and $f \circ H_t^{n+1} = H_t^n \circ f$ for all $t \in I$ and $n \in \mathbb{N}_0$.

Claim: $H_1^n(\mathcal{C}) = \mathcal{C}'$ for $n \in \mathbb{N}_0$. To see this we use induction on n . For $n = 0$ the claim is true by choice of H^0 .

Suppose that $H_1^n(\mathcal{C}) = \mathcal{C}'$ for some $n \in \mathbb{N}_0$. Then Lemma 10.2 and the identity $f \circ H_1^{n+1} = H_1^n \circ f$ imply that

$$H_1^{n+1}(f^{-1}(\mathcal{C})) = f^{-1}(H_1^n(\mathcal{C})) = f^{-1}(\mathcal{C}').$$

Since \mathcal{C} and \mathcal{C}' are f -invariant, we have the inclusions $\mathcal{C} \subset f^{-1}(\mathcal{C})$ and $\mathcal{C}' \subset f^{-1}(\mathcal{C}')$. In particular,

$$\tilde{\mathcal{C}} := H_1^{n+1}(\mathcal{C}) \subset H_1^{n+1}(f^{-1}(\mathcal{C})) = f^{-1}(\mathcal{C}')$$

is a Jordan curve contained in $f^{-1}(\mathcal{C}')$. Moreover, \mathcal{C} and $\tilde{\mathcal{C}}$ are isotopic rel. $f^{-1}(\text{post}(f))$ (by the isotopy H^{n+1}). Since \mathcal{C} and \mathcal{C}' are isotopic rel. $f^{-1}(\text{post}(f))$ by our hypotheses, it follows that \mathcal{C}' and $\tilde{\mathcal{C}}$ are also isotopic rel. $f^{-1}(\text{post}(f))$. Both sets are contained in $f^{-1}(\mathcal{C}')$.

Now $f^{-1}(\mathcal{C}')$ is the 1-skeleton of the cell decomposition $\mathcal{D}^1(f, \mathcal{C}')$. This cell decomposition has the vertex set $f^{-1}(\text{post}(f))$. Moreover, since f is expanding, $\text{post}(f) \geq 3$, and so every tile in $\mathcal{D}^1(f, \mathcal{C}')$ has at least three vertices. So the hypotheses of Lemma 10.12 are satisfied and we conclude that $\mathcal{C}' = \tilde{\mathcal{C}} = H_1^{n+1}(\mathcal{C})$. The claim follows.

Fix a visual metric on S^2 . Then the tracks of the isotopies H^n shrink at an exponential rate as $n \rightarrow \infty$ (Lemma 10.3). Since $H_0^n = \text{id}_{S^2}$, it follows that $H_1^n \rightarrow \text{id}_{S^2}$ uniformly as $n \rightarrow \infty$. Since $H_1^n(\mathcal{C}) = \mathcal{C}'$ for all $n \in \mathbb{N}_0$ by the claim, we conclude $\mathcal{C} = \mathcal{C}'$ as desired. \square

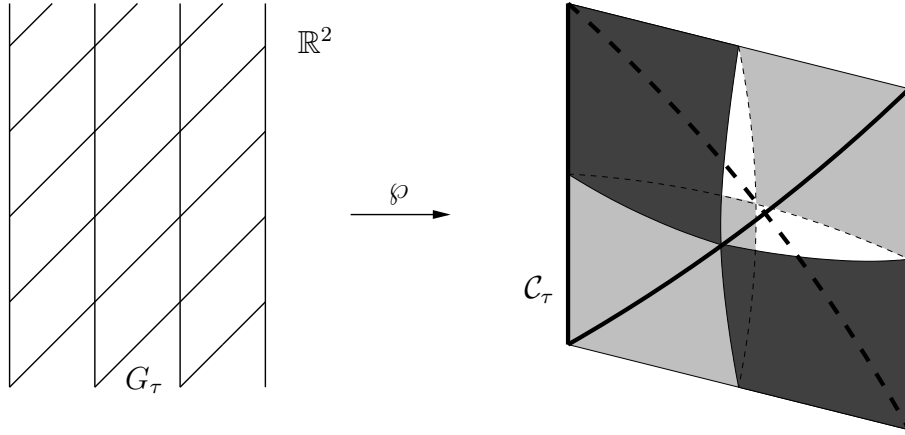
Corollary 15.2 (Inv. curves in a given isotopy class rel. $\text{post}(f)$).

Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map and $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$. Then there are at most finitely many f -invariant Jordan curves $\tilde{\mathcal{C}} \subset S^2$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$ that are isotopic to \mathcal{C} rel. $\text{post}(f)$.

Proof. Let $\tilde{\mathcal{C}}$ be such an f -invariant Jordan curve. Then there exists an isotopy $H: S^2 \times I \rightarrow S^2$ rel. $\text{post}(f)$ with $H_0 = \text{id}_{S^2}$ and $H_1(\tilde{\mathcal{C}}) = \mathcal{C}$. Lifting H we get an isotopy $\tilde{H}: S^2 \times I \rightarrow S^2$ rel. $f^{-1}(\text{post}(f))$ such that $\tilde{H}_0 = \text{id}_{S^2}$ and $f \circ \tilde{H}_t = H_t \circ f$ for all $t \in I$. Since $\tilde{\mathcal{C}}$ is f -invariant, we have $\tilde{\mathcal{C}} \subset f^{-1}(\tilde{\mathcal{C}})$. So Lemma 10.2 implies that

$$\tilde{H}_1(\tilde{\mathcal{C}}) \subset \tilde{H}_1(f^{-1}(\tilde{\mathcal{C}})) = f^{-1}(\tilde{H}_1(\tilde{\mathcal{C}})) = f^{-1}(\mathcal{C}).$$

Hence $\tilde{\mathcal{C}}$ is isotopic rel. $f^{-1}(\text{post}(f))$ to the Jordan curve $\tilde{H}_1(\tilde{\mathcal{C}})$ that is contained in $f^{-1}(\mathcal{C})$. Any such Jordan curve is a union of edges in the cell decomposition $\mathcal{D}^1(f, \mathcal{C})$ (see the last part of the proof of Lemma 10.12). In particular, there are only finitely many distinct Jordan curves contained in $f^{-1}(\mathcal{C})$. This implies that there are only finitely many isotopy classes rel. $f^{-1}(\text{post}(f))$ represented by curves $\tilde{\mathcal{C}}$ satisfying the assumptions of the corollary. Since an f -invariant Jordan curve $\tilde{\mathcal{C}} \subset S^2$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$ is unique in its isotopy class rel. $f^{-1}(\text{post}(f))$ by Theorem 1.4, the statement follows. \square

FIGURE 19. Invariant curves for the Lattès map g .

Corollary 15.3. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map with $\text{post}(f) = 3$. Then there are at most finitely many f -invariant Jordan curves $\tilde{\mathcal{C}} \subset S^2$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$.*

Proof. Pick a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$. Since we have $\#\text{post}(f) = 3$, by Lemma 10.10 every Jordan curve $\tilde{\mathcal{C}} \subset S^2$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$ is isotopic to \mathcal{C} rel. $\text{post}(f)$. The statement now follows from Corollary 15.2. \square

In contrast to the case $\#\text{post}(f) = 3$, expanding Thurston maps f with $\#\text{post}(f) \geq 4$ can have infinitely many distinct invariant Jordan curves containing the set of postcritical points.

Example 15.4 (Infinitely many invariant curves). Let g be the Lattès map from Section 1.2. In the following it is advantageous to use real notation and consider the maps ψ and \wp used in the definition of g as in (1.2) as maps on \mathbb{R}^2 . Then $\psi(u) = 2u$ for $u \in \mathbb{R}^2$. Moreover, for $u_1, u_2 \in \mathbb{R}^2$ we have $\wp(u_1) = \wp(u_2)$ if and only if $u_2 = \pm u_1 + \gamma$ for $\gamma \in L = 2\mathbb{Z}^2$. Recall that the extended real line $\hat{\mathbb{R}}$ (which is the boundary of the pillow) is g -invariant and contains $\{-1, 0, 1, \infty\} = \text{post}(g)$. Let $S = \partial Q$ be the boundary of the unit square $Q = [0, 1]^2 \subset \mathbb{R}^2$, and G be the standard square grid, i.e., the union of the horizontal and vertical lines through the points in \mathbb{Z}^2 . Then $\wp|_S$ is injective and $\wp(S) = \wp(G) = \hat{\mathbb{R}}$. So the g -invariant curve $\hat{\mathbb{R}}$ is obtained by mapping the boundary S of the fundamental domain Q of \mathbb{Z}^2 or the standard grid G by \wp . One can obtain other g -invariant Jordan curves by mapping the boundaries of other fundamental domains of \mathbb{Z}^2 or other grids by \wp .

To explain this, consider a (2×2) -matrix $\tau \in \mathrm{SL}_2(\mathbb{Z})$. We identify τ with the linear map $u \mapsto \tau u$ on \mathbb{R}^2 induced by left-multiplication of $u \in \mathbb{R}^2$ (considered as a column vector) by the matrix τ . Define $Q_\tau := \tau(Q)$, $S_\tau := \partial Q_\tau = \tau(S)$, and the corresponding grid $G_\tau = \tau(G)$. Since τ is a linear map and $\tau(L) = L$, it follows that for $u_1, u_2 \in \mathbb{R}^2$ we have $\wp(\tau(u_1)) = \wp(\tau(u_2))$ if and only if $u_2 = \pm u_1 + \gamma$, where $\gamma \in L$. This implies that $\wp|_{S_\tau}$ is injective, and so $\mathcal{C}_\tau := \wp(S_\tau) \subset \widehat{\mathbb{C}}$ is a Jordan curve. Moreover, $\wp(G_\tau) = \wp(S_\tau)$. Since $\psi \circ \tau = \tau \circ \psi$, we have

$$\psi(G_\tau) = \psi(\tau(G)) = \tau(\psi(G)) \subset \tau(G) = G_\tau,$$

and so

$$g(\mathcal{C}_\tau) = g(\wp(G_\tau)) = \wp(\psi(G_\tau)) \subset \wp(G_\tau) = \mathcal{C}_\tau.$$

Hence \mathcal{C}_τ is g -invariant. Since $\mathbb{Z}^2 \subset G_\tau$, we also have $\mathrm{post}(g) = \wp(\mathbb{Z}^2) \subset \wp(G_\tau) = \mathcal{C}_\tau$. So \mathcal{C}_τ is an g -invariant Jordan curve that contains the set $\mathrm{post}(g)$. An example of this construction is indicated in Figure 19. The curve \mathcal{C}_τ is drawn in thick on the right.

The curve \mathcal{C}_τ determines the grid G_τ uniquely; indeed, one obtains generating vectors of the two lines in G_τ through 0 by locally lifting \mathcal{C}_τ near $\wp(0) = 0 \in \mathrm{post}(g) \subset \mathcal{C}_\tau$ to 0 by the map \wp . The whole grid G_τ is obtained by translating these two lines by vectors in \mathbb{Z}^2 .

This implies that the map $\tau \in \mathrm{SL}_2(\mathbb{Z}) \mapsto \mathcal{C}_\tau$ is four-to-one; indeed, if $\tau, \sigma \in \mathrm{SL}_2(\mathbb{Z})$, then, as we have seen, $\mathcal{C}_\tau = \mathcal{C}_\sigma$ if and only if $G_\tau = G_\sigma$. On the other hand, $G_\tau = G_\sigma$ if and only if $\sigma^{-1} \circ \tau \in \mathrm{SL}_2(\mathbb{Z})$ is one of the four rotations (by integer multiples of $\pi/2$) that preserve the grid G . In particular, there exist infinitely many g -invariant Jordan curves $\tilde{\mathcal{C}} \subset \widehat{\mathbb{C}}$ with $\mathrm{post}(g) \subset \tilde{\mathcal{C}}$.

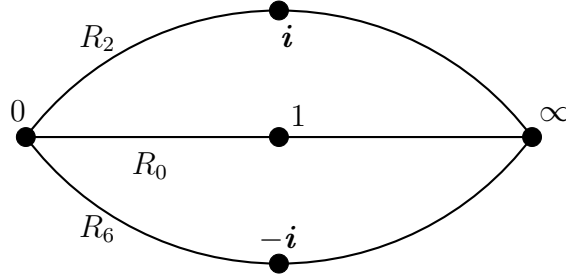
We now turn to existence results. As the following example shows, for a Thurston map $f: S^2 \rightarrow S^2$ an f -invariant Jordan curve $\mathcal{C} \subset S^2$ with $\mathrm{post}(f) \subset \mathcal{C}$ need not exist.

Example 15.5. Consider the map $f: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ defined by

$$f(z) = i \frac{z^4 - i}{z^4 + i}$$

for $z \in \widehat{\mathbb{C}}$. The critical points of f are 0 and ∞ . The forward orbits of the critical points of f under iteration can be represented by the so-called *ramification portrait*:

$$(15.1) \quad \begin{array}{ccc} 0 & \xrightarrow{4:1} & -i \\ \infty & \xrightarrow{4:1} & i \end{array} \quad \begin{array}{c} \nearrow \\ \searrow \end{array} \quad \begin{array}{c} \curvearrowright \\ 1 \end{array}$$

FIGURE 20. No invariant Jordan curve $\tilde{\mathcal{C}} \supset \text{post}$.

The labels over the arrows indicate the local degree if it is different from 1. It follows that the set of postcritical points of f is given by $\text{post}(f) = \{-i, 1, i\}$. So f is a Thurston map. This map is also expanding as follows from Proposition 19.1 below.

Lemma 15.6. *Let f be the map from Example 15.5. Then there is no f -invariant Jordan curve $\tilde{\mathcal{C}} \subset \hat{\mathbb{C}}$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$.*

Proof. We have $f(z) = \varphi(z^4)$ for $z \in \hat{\mathbb{C}}$, where

$$(15.2) \quad \varphi(w) = i \frac{w - i}{w + i}, \quad w \in \hat{\mathbb{C}},$$

is a Möbius transformation that maps the upper half-plane to the unit disk (note that φ maps $0, 1, \infty$ to $-i, 1, i$, respectively). Let $\mathcal{C} := \partial\mathbb{D}$ be the unit circle. Then

$$f^{-1}(\mathcal{C}) = \bigcup_{k=0, \dots, 7} R_k, \quad \text{where } R_k = \{re^{ik\pi/4} : 0 \leq r \leq \infty\}.$$

The postcritical points $-i, 1, i$ lie on distinct rays R_k . Two such rays have the points 0 and ∞ in common and no other points. Thus there is no Jordan curve in $f^{-1}(\mathcal{C})$ containing all postcritical points, see Figure 20. As we will see in Theorem 1.3, the existence of such a Jordan curve is a necessary condition for the existence of an f -invariant Jordan curve $\tilde{\mathcal{C}} \subset \hat{\mathbb{C}}$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$ (in our specific case where $\#\text{post}(f) = 3$ the choice of \mathcal{C} does not matter since all Jordan curves that contain $\text{post}(f)$ are isotopic rel. $\text{post}(f)$). Hence there is no f -invariant Jordan curve $\tilde{\mathcal{C}} \subset \hat{\mathbb{C}}$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$. One can also see this by a simple argument directly.

Indeed, suppose that $\tilde{\mathcal{C}} \subset \hat{\mathbb{C}}$ is a Jordan curve with $\text{post}(f) \subset \tilde{\mathcal{C}}$ and $f(\tilde{\mathcal{C}}) \subset \tilde{\mathcal{C}}$. The unit circle $\mathcal{C} = \partial\mathbb{D}$ is also a Jordan curve containing the set $\text{post}(f)$. Hence by Lemma 10.10 there exists an isotopy $H: \hat{\mathbb{C}} \times I \rightarrow \hat{\mathbb{C}}$ rel. $\text{post}(f)$ such that $H_0 = \text{id}_{\hat{\mathbb{C}}}$ and $H_1(\tilde{\mathcal{C}}) = \mathcal{C}$.

By Proposition 10.1 the isotopy H can be lifted to an isotopy $\tilde{H}: \widehat{\mathbb{C}} \times I \rightarrow \widehat{\mathbb{C}}$ rel. $\text{post}(f)$ such that $\tilde{H}_0 = \text{id}_{\widehat{\mathbb{C}}}$ and $H_t \circ f = f \circ \tilde{H}_t$ for all $t \in I$.

Since $\tilde{\mathcal{C}} \subset f^{-1}(\tilde{\mathcal{C}})$, it follows from Lemma 10.2 that

$$\tilde{H}_1(\tilde{\mathcal{C}}) \subset \tilde{H}_1(f^{-1}(\tilde{\mathcal{C}})) = f^{-1}(H_1(\tilde{\mathcal{C}})) = f^{-1}(\mathcal{C}).$$

This means that the Jordan curve $\mathcal{C}' := \tilde{H}_1(\tilde{\mathcal{C}})$ is contained in $f^{-1}(\mathcal{C})$. Moreover, it contains all postcritical points, since $\tilde{\mathcal{C}}$ does, and the points in $\text{post}(f)$ stay fixed under the isotopy \tilde{H} . As we have seen above, no such Jordan curve exists and we get a contradiction as desired. \square

By a similar (though somewhat more lengthy) argument one can show that the Lattès map $f(z) = \frac{i}{2}(z + 1/z)$ does not have an f -invariant curve \mathcal{C} with $\text{post}(f) \subset \mathcal{C}$. Another such example can be found in [CFP10, Section 4].

We now turn to the proof for the necessary and sufficient criterion for the existence of an invariant Jordan curve as formulated in Theorem 1.3. Note that in (ii) of this theorem the condition on \hat{f} is meaningful. Indeed, $\hat{f} = H_1 \circ f$ is a branched cover of S^2 with $\text{crit}(\hat{f}) = \text{crit}(f)$. Since H_1 is isotopic to id_{S^2} rel. $\text{post}(f)$ we have $\text{post}(\hat{f}) = \text{post}(f)$, and so \hat{f} is a Thurston map.

Furthermore, \mathcal{C}' is a Jordan curve with $\text{post}(\hat{f}) = \text{post}(f) \subset \mathcal{C}'$. Since $\mathcal{C}' = H_1(\mathcal{C}) \subset f^{-1}(\mathcal{C})$, we have that $\hat{f}(\mathcal{C}') = H_1 \circ f(\mathcal{C}') \subset H_1(\mathcal{C}) = \mathcal{C}'$. Hence \mathcal{C}' is invariant with respect to \hat{f} , and it makes sense to require that \hat{f} is combinatorially expanding for \mathcal{C}' .

Proof of Theorem 1.3. (i) \Rightarrow (ii): Suppose $\tilde{\mathcal{C}}$ is as in (i). Then in (ii) we let $\mathcal{C} = \mathcal{C}' = \tilde{\mathcal{C}}$, and the isotopy H be such that $H_t = \text{id}_{S^2}$ for all $t \in I$. Then $\mathcal{C}' = \tilde{\mathcal{C}} \subset f^{-1}(\tilde{\mathcal{C}}) = f^{-1}(\mathcal{C})$, and $\hat{f} = f$ is combinatorially expanding for the invariant curve $\mathcal{C}' = \tilde{\mathcal{C}}$, since f is expanding.

(ii) \Rightarrow (i): Let $\mathcal{C}, \mathcal{C}', H$, and \hat{f} be as in (ii), and define $\chi = H_1$. As we have seen in the remark after the theorem, \hat{f} is a Thurston map with $\text{post}(\hat{f}) = \text{post}(f)$, and \mathcal{C}' is an \hat{f} -invariant Jordan curve containing the set $\text{post}(\hat{f}) = \text{post}(f)$.

Since \hat{f} is combinatorially expanding for \mathcal{C}' , Corollary 13.18 implies that there exists a homeomorphism $\phi: S^2 \rightarrow S^2$ that is isotopic to the identity rel. $\text{post}(\hat{f}) = \text{post}(f)$ such that $\phi(\mathcal{C}') = \mathcal{C}'$ and $g = \phi \circ \hat{f}$ is an expanding Thurston map. Since $g = (\phi \circ \chi) \circ f$, and $\phi \circ \chi$ is isotopic to the identity on S^2 rel. $\text{post}(f)$, the maps f and g are Thurston equivalent. If notation is as in (3.3) (with $\widehat{S^2} = S^2$), then we can take $h_0 = \phi \circ \chi$ and $h_1 = \text{id}_{S^2}$. By Theorem 10.4 we

can find a homeomorphism $h: S^2 \rightarrow S^2$ that is isotopic to $h_1 = \text{id}_{S^2}$ rel. $f^{-1}(\text{post}(f))$ with $f \circ h = h \circ g$.

Let $\tilde{\mathcal{C}} = h(\mathcal{C}')$. Then $\tilde{\mathcal{C}}$ is a Jordan curve in S^2 that is isotopic to \mathcal{C}' rel. $f^{-1}(\text{post}(f))$, and hence isotopic to \mathcal{C} rel. $\text{post}(f)$; in particular, $\tilde{\mathcal{C}}$ contains the set $\text{post}(f)$. Moreover, $\tilde{\mathcal{C}}$ is f -invariant, because we have

$$\begin{aligned} f(\tilde{\mathcal{C}}) &= f(h(\mathcal{C}')) = h(g(\mathcal{C}')) = h(\phi(\hat{f}(\mathcal{C}'))) \\ &\subset h(\phi(\mathcal{C}')) = h(\mathcal{C}') = \tilde{\mathcal{C}}. \end{aligned}$$

The proof is complete. \square

Remarks 15.7. (a) The condition of combinatorial expansion in (ii) of Theorem 1.3 is combinatorial in nature and can easily be checked in principle. A simple sufficient criterion for this can be formulated as follows: if no 1-tile for (f, \mathcal{C}) joins opposite sides of \mathcal{C}' , then \hat{f} is combinatorially expanding for \mathcal{C}' .

To see this note that

$$\hat{f}^{-1}(\mathcal{C}') = f^{-1}(H_1^{-1}(\mathcal{C}')) = f^{-1}(\mathcal{C}).$$

By Proposition 6.1 (v) this implies that the 1-tiles for (\hat{f}, \mathcal{C}') are precisely the 1-tiles for (f, \mathcal{C}) . Hence if no 1-tile for (f, \mathcal{C}) joins opposite sides of \mathcal{C}' , then $D_1(\hat{f}, \mathcal{C}') \geq 2$ and so \hat{f} is combinatorially expanding for \mathcal{C}' . We will later formulate a necessary and sufficient condition for combinatorial expansion of \hat{f} (see Proposition 15.13).

(b) The condition of combinatorial expansion in (ii) is independent of the chosen isotopy H . More precisely, suppose $H^1, H^2: S^2 \times I \rightarrow S^2$ are two isotopies with $H_0^1 = H_0^2 = \text{id}_{S^2}$ and $H_1^1(\mathcal{C}) = H_1^2(\mathcal{C}) = \mathcal{C}'$. Then $\hat{f}_1 = H_1^1 \circ f$ is combinatorially expanding for \mathcal{C}' if and only if $\hat{f}_2 = H_1^2 \circ f$ is combinatorially expanding for \mathcal{C}' . This follows immediately from Lemma 12.8 (with $f = \hat{f}_1$, $g = \hat{f}_2$, $h_0 = H_1^2 \circ (H_1^1)^{-1}$, $h_1 = \text{id}_{S^2}$, and $\mathcal{C} = \mathcal{C}'$).

(c) Theorem 1.3 can be slightly modified to give necessary and sufficient conditions for the existence of an invariant curve in a given isotopy class rel. $\text{post}(f)$ or rel. $f^{-1}(\text{post}(f))$. An existence statement for a given isotopy class rel. $f^{-1}(\text{post}(f))$ is especially relevant in view of the complementary uniqueness statement given by Theorem 1.4.

To formulate this precisely, let $\hat{\mathcal{C}} \subset S^2$ be given a Jordan curve with $\text{post}(f) \subset \hat{\mathcal{C}}$. Then an f -invariant Jordan curve $\tilde{\mathcal{C}} \subset S^2$ isotopic to $\hat{\mathcal{C}}$ rel. $\text{post}(f)$ exists if and only if condition (ii) in Theorem 1.3 is true for a Jordan curve \mathcal{C} isotopic to $\hat{\mathcal{C}}$ rel. $\text{post}(f)$. This immediately follows from the proof of this theorem.

Similarly, an f -invariant Jordan curve $\tilde{\mathcal{C}} \subset S^2$ isotopic to $\hat{\mathcal{C}}$ rel. $f^{-1}(\text{post}(f))$ exists if and only if condition (ii) in Theorem 1.3 is true with the additional requirement that \mathcal{C}' is isotopic to $\hat{\mathcal{C}}$ rel. $f^{-1}(\text{post}(f))$.

Existence of an f -invariant Jordan curve $\tilde{\mathcal{C}}$ in a given isotopy class rel. $f^{-1}(\text{post}(f))$ is particularly interesting, because if there is such a curve $\tilde{\mathcal{C}}$, then it is unique by Theorem 1.4. It is worthwhile to formulate some explicit conditions which guarantee existence in this. They are stated in the next proposition and are a slight variation of the condition given in Remark 15.7 (c).

Proposition 15.8 (Existence of invariant curves rel. $f^{-1}(\text{post}(f))$).

Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map and $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$. Then the following conditions are equivalent:

- (i) There exists an f -invariant Jordan curve $\tilde{\mathcal{C}}$ that is isotopic to \mathcal{C} rel. $f^{-1}(\text{post}(f))$.
- (ii) There exists an isotopy $H: S^2 \times I \rightarrow S^2$ rel. $f^{-1}(\text{post}(f))$ with $H_0 = \text{id}_{S^2}$ and $H_1(\mathcal{C}) \subset f^{-1}(\mathcal{C})$ such that

$$\hat{g} := f \circ H_1$$

is combinatorially expanding for \mathcal{C} .

- (iii) There exists an isotopy $H: S^2 \times I \rightarrow S^2$ rel. $f^{-1}(\text{post}(f))$ with $H_0 = \text{id}_{S^2}$ and $\mathcal{C}' := H_1(\mathcal{C}) \subset f^{-1}(\mathcal{C})$ such that

$$\hat{f} := H_1 \circ f$$

is combinatorially expanding for \mathcal{C}' .

Proof. (i) \Rightarrow (ii): Suppose that there exists an f -invariant Jordan $\tilde{\mathcal{C}}$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$ that is isotopic to \mathcal{C} rel. $f^{-1}(\text{post}(f))$. Then we can find an isotopy $K: S^2 \times I \rightarrow S^2$ rel. $f^{-1}(\text{post}(f))$ with $K_0 = \text{id}_{S^2}$ and $K_1(\tilde{\mathcal{C}}) = \mathcal{C}$.

Let \tilde{K} be the lift of K by f according to Proposition 10.1, i.e., the isotopy rel. $f^{-2}(\text{post}(f)) \supset f^{-1}(\text{post}(f))$ with $\tilde{K}_0 = \text{id}_{S^2}$ and $K_t \circ f = f \circ \tilde{K}_t$ for $t \in I$. Define an isotopy $H: S^2 \times I \rightarrow S^2$ by setting $H_t = \tilde{K}_t \circ (K_t)^{-1}$ for $t \in I$. Then H is an isotopy rel. $f^{-1}(\text{post}(f))$ with $H_0 = \text{id}_{S^2}$. Moreover,

$$\begin{aligned} H_1(\mathcal{C}) &= (\tilde{K}_1 \circ (K_1)^{-1})(\mathcal{C}) = \tilde{K}_1(\tilde{\mathcal{C}}) \subset \tilde{K}_1(f^{-1}(\tilde{\mathcal{C}})) \\ &= f^{-1}(K_1(\tilde{\mathcal{C}})) = f^{-1}(\mathcal{C}). \end{aligned}$$

Here we used the f -invariance of $\tilde{\mathcal{C}}$ (i.e., $\tilde{\mathcal{C}} \subset f^{-1}(\tilde{\mathcal{C}})$), as well as Lemma 10.2. Note that

$$\hat{g} := f \circ H_1 = f \circ \tilde{K}_1 \circ (K_1)^{-1} = K_1 \circ f \circ (K_1)^{-1}.$$

Hence \hat{g} is a Thurston map with $\text{post}(\hat{g}) = \text{post}(f)$, \mathcal{C} is an g -invariant Jordan curve with $\mathcal{C} \supset \text{post}(g)$, and it follows from Lemma 12.8 that \hat{g} is combinatorially expanding for \mathcal{C} .

(ii) \Rightarrow (iii): Let H and \hat{g} be as in (ii), and define $\hat{f} := H_1 \circ f$ and $\mathcal{C}' := H_1(\mathcal{C})$. Then \hat{f} is a Thurston map with $\text{post}(\hat{f}) = \text{post}(f)$, and we have

$$\hat{f}(\mathcal{C}') = \hat{f}(H_1(\mathcal{C})) \subset \hat{f}(f^{-1}(\mathcal{C})) = H_1(f(f^{-1}(\mathcal{C}))) = H_1(\mathcal{C}) = \mathcal{C}'.$$

Hence \mathcal{C}' is an \hat{f} -invariant Jordan curve with $\text{post}(\hat{f}) \subset \mathcal{C}'$. Since

$$\hat{f} \circ H_1 = H_1 \circ f \circ H_1 = H_1 \circ \hat{g},$$

Lemma 12.8 implies that \hat{f} is combinatorially expanding for \mathcal{C}' .

(iii) \Rightarrow (i): If (iii) is true, then by Theorem 1.3 there exists an f -invariant Jordan curve $\tilde{\mathcal{C}}$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$ that is isotopic to \mathcal{C}' rel. $f^{-1}(\text{post}(f))$. Since $\mathcal{C}' = H_1(\mathcal{C})$ and H is an isotopy rel. $f^{-1}(\text{post}(f))$, the curve \mathcal{C}' , and hence also $\tilde{\mathcal{C}}$, is isotopic to \mathcal{C} rel. $f^{-1}(\text{post}(f))$. \square

We now turn to the proof of Theorem 1.2. We need the following auxiliary result.

Lemma 15.9. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$. Then for all sufficiently large n there exists a Jordan curve $\mathcal{C}' \subset f^{-n}(\mathcal{C})$ that is isotopic to \mathcal{C} rel. $\text{post}(f)$. Moreover, \mathcal{C}' can be chosen so that no n -tile for (f, \mathcal{C}) joins opposite sides of \mathcal{C}' .*

Proof. We fix some base metric on S^2 . Let $P := \text{post}(f)$. Since f is expanding, we have $k := \#P = \#\text{post}(f) \geq 3$ by Corollary 6.4. Pick $\epsilon_0 > 0$ as in Lemma 14.5. Since f is expanding, for large enough n we have

$$\text{mesh}(f, n, \mathcal{C}) = \max_{c \in \mathcal{D}^n(f, \mathcal{C})} \text{diam}(c) < \epsilon_0.$$

For such n consider the cell decomposition $\mathcal{D} = \mathcal{D}^n(f, \mathcal{C})$ of S^2 . Its vertex set is the set $f^{-n}(\text{post}(f)) \supset \text{post}(f) = P$ of n -vertices and its 1-skeleton is the set $f^{-n}(\mathcal{C})$. Hence by Lemma 14.5 there exists a Jordan curve $\mathcal{C}' \subset f^{-n}(\mathcal{C})$ that is isotopic to \mathcal{C} rel. $P = \text{post}(f)$ and so that no tile in \mathcal{D} , i.e., no n -tile for (f, \mathcal{C}) , joins opposite side of \mathcal{C}' . \square

Proof of Theorem 1.2. Let f and \mathcal{C} be as in the statement of the theorem. By Lemma 15.9 for sufficiently large $n \in \mathbb{N}$ there exists an isotopy $H: S^2 \times I \rightarrow S^2$ rel. $\text{post}(f)$ such that $H_0 = \text{id}_{S^2}$ and $\mathcal{C}' := H_1(\mathcal{C}) \subset f^{-n}(\mathcal{C})$ and such that no n -tile for (f, \mathcal{C}) joins opposite sides of \mathcal{C}' .

For such n let $F = f^n$. Then F is an expanding Thurston map with $\text{post}(F) = \text{post}(f)$. Then \mathcal{C} and \mathcal{C}' are Jordan curves with $\text{post}(F) \subset \mathcal{C}, \mathcal{C}'$, and H is an isotopy rel. $\text{post}(F)$ that deforms \mathcal{C} into $\mathcal{C}' \subset f^{-n}(\mathcal{C}) = F^{-1}(\mathcal{C})$. Moreover, $F = f^n$ is cellular for $(\mathcal{D}^n(f, \mathcal{C}), \mathcal{D}^0(f, \mathcal{C}))$. Hence by Lemma 5.4 we have $\mathcal{D}^1(F, \mathcal{C}) = \mathcal{D}^n(f, \mathcal{C})$, and so the 1-cells for (F, \mathcal{C}) are precisely the n -cells for (f, \mathcal{C}) . So no 1-tile for (F, \mathcal{C}) joins opposite side of \mathcal{C}' and by Remark 15.7 (a) the map $H_1 \circ F$ is combinatorially expanding for \mathcal{C}' . This shows that condition (ii) in Theorem 1.3 is satisfied. Hence there exists a Jordan curve $\tilde{\mathcal{C}} \subset S^2$ that is F -invariant and is isotopic to \mathcal{C} rel. $\text{post}(F) = \text{post}(f)$ as desired. \square

Remark 15.10. In general, the f^n -invariant Jordan curve $\tilde{\mathcal{C}}$ as in Theorem 1.2 will depend on n , and one cannot expect that $\tilde{\mathcal{C}}$ is invariant for *all* sufficiently high iterates of f . To illustrate this, consider the map f from Example 15.5 (see also Lemma 15.6). Recall that $f(z) = \varphi(z^4)$ for $z \in \hat{\mathbb{C}}$, where φ is as in (15.2).

The Möbius transformation φ maps the extended real line $\hat{\mathbb{R}}$ to the unit circle $\partial\mathbb{D}$, and $\partial\mathbb{D}$ to $\hat{\mathbb{R}}$. This implies the unit circle $\tilde{\mathcal{C}} := \partial\mathbb{D}$ satisfies $f^{2n}(\tilde{\mathcal{C}}) \subset \tilde{\mathcal{C}}$ for every $n \in \mathbb{N}$. Note that $\text{post}(f) = \{-i, 1, i\} \subset \tilde{\mathcal{C}}$. Thus $\tilde{\mathcal{C}}$ is a Jordan curve with $\text{post}(f) \subset \tilde{\mathcal{C}}$ that is invariant for every *even* iterate f^{2n} .

On the other hand, for $n \in \mathbb{N}_0$ we have $f^{2n+1}(\partial\mathbb{D}) \subset \hat{\mathbb{R}}$. Since f^{2n+1} is a finite-to-one map, the set $f^{2n+1}(\partial\mathbb{D})$ is infinite, and so we cannot have $f^{2n+1}(\partial\mathbb{D}) \subset \partial\mathbb{D}$ (for otherwise, $f^{2n+1}(\partial\mathbb{D}) \subset \partial\mathbb{D} \cap \hat{\mathbb{R}} = \{-1, 1\}$). Thus the unit circle $\partial\mathbb{D} = \tilde{\mathcal{C}}$ is not invariant for any *odd* iterate of f .

Proof of Corollary 1.5. Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map. It follows from Theorem 1.2 that for each sufficiently large $n \in \mathbb{N}$ there exists an f^n -invariant Jordan curve $\tilde{\mathcal{C}} \subset S^2$ with $\text{post}(f) = \text{post}(f^n) \subset \tilde{\mathcal{C}}$. For such n let $F = f^n$, $\mathcal{D}^0 = \mathcal{D}^0(F, \tilde{\mathcal{C}})$, and $\mathcal{D}^1(F, \mathcal{C})$. Define the orientation-preserving labeling $L: \mathcal{D}^1 \rightarrow \mathcal{D}^0$ by $L(c) = F(c)$ for $c \in \mathcal{D}^1$. Then it is clear that $(\mathcal{D}^1, \mathcal{D}^0, L)$ is a two-tile subdivision rule that is realized by F (see Definition 12.4 and the following discussion). \square

15.2. Iterative construction of invariant curves.

Given data as in Theorem 1.3 (ii), the f -invariant curve $\tilde{\mathcal{C}}$ can be obtained by an iterative procedure. To explain this, we first recall the definition of *Hausdorff convergence* of sets.

Let (X, d) be a compact metric space. If $A, B \subset X$ are subsets of X , then their *Hausdorff distance* is defined as

$$(15.3) \quad \text{dist}_d^H(A, B) = \inf\{\delta > 0 : A \subset \mathcal{N}_d^\delta(B) \text{ and } B \subset \mathcal{N}_d^\delta(A)\}.$$

Assume A and A_n for $n \in \mathbb{N}$ are closed subsets of X . We say that $A_n \rightarrow A$ as $n \rightarrow \infty$ in the sense of *Hausdorff convergence* if

$$\lim_{n \rightarrow \infty} \text{dist}_d^H(A_n, A) = 0.$$

Note that in this case a point $x \in X$ lies in A if and only if there exists a sequence (x_n) of points in X such that $x_n \in A_n$ for $n \in \mathbb{N}$ and $x_n \rightarrow x$ as $n \rightarrow \infty$.

Now let $f: S^2 \rightarrow S^2$ be an expanding Thurston map, $\mathcal{C}, \mathcal{C}' \subset S^2$ be Jordan curves with $\text{post}(f) \subset \mathcal{C}, \mathcal{C}'$ and $\mathcal{C}' \subset f^{-1}(\mathcal{C})$, and let $H: S^2 \times I \rightarrow S^2$ be an isotopy rel. $\text{post}(f)$ that deforms \mathcal{C} to \mathcal{C}' , i.e., $H_0 = \text{id}_{S^2}$ and $H_1(\mathcal{C}) = \mathcal{C}'$. For the moment we do *not* assume that the map $\hat{f} = H_1 \circ f$ is combinatorially expanding for \mathcal{C}' .

Let $H^0 := H$. Using Proposition 10.1 repeatedly, we can find isotopies $H^n: S^2 \times I \rightarrow S^2$ rel. $f^{-1}(\text{post}(f))$ such that $H_0^n = \text{id}_{S^2}$ and $f \circ H_t^n = H_t^{n-1} \circ f$ for all $n \in \mathbb{N}$, $t \in I$. Now define Jordan curves inductively by setting $\mathcal{C}^0 := \mathcal{C}$, and $\mathcal{C}^{n+1} := H_1^n(\mathcal{C}^n)$ for $n \in \mathbb{N}_0$. Note that then $\mathcal{C}^1 = \mathcal{C}'$.

To summarize, we start with the following data (for a given f):

- a Jordan curve $\mathcal{C}^0 = \mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}^0$,
- a Jordan curve $\mathcal{C}^1 = \mathcal{C}' \subset S^2$ isotopic to $\mathcal{C}^0 \subset S^2$ rel. $\text{post}(f)$ with $\mathcal{C}^1 \subset f^{-1}(\mathcal{C}^0)$,
- an isotopy $H^0: S^2 \times I \rightarrow S^2$ rel. $\text{post}(f)$ such that $H_0^0 = \text{id}_{S^2}$ and $H_1^0(\mathcal{C}^0) = \mathcal{C}^1$.

We then define inductively:

- isotopies $H^n: S^2 \times I \rightarrow S^2$ such that $H_0^n = \text{id}_{S^2}$ and $f \circ H_t^n = H_t^{n-1} \circ f$ for all $n \in \mathbb{N}$, $t \in I$,
- Jordan curves $\mathcal{C}^{n+1} := H_1^n(\mathcal{C}^n)$ for $n \in \mathbb{N}_0$.

Figure 18 illustrates this procedure for Example 15.1. Since this example is rather complicated and it is hard to grasp the involved isotopies, we present a simpler example for the construction.

Example 15.11. Let f be the Lattès map obtained as in (1.2), where we choose

$$\psi: \mathbb{C} \rightarrow \mathbb{C}, \quad u \mapsto \psi(u) := 5u.$$

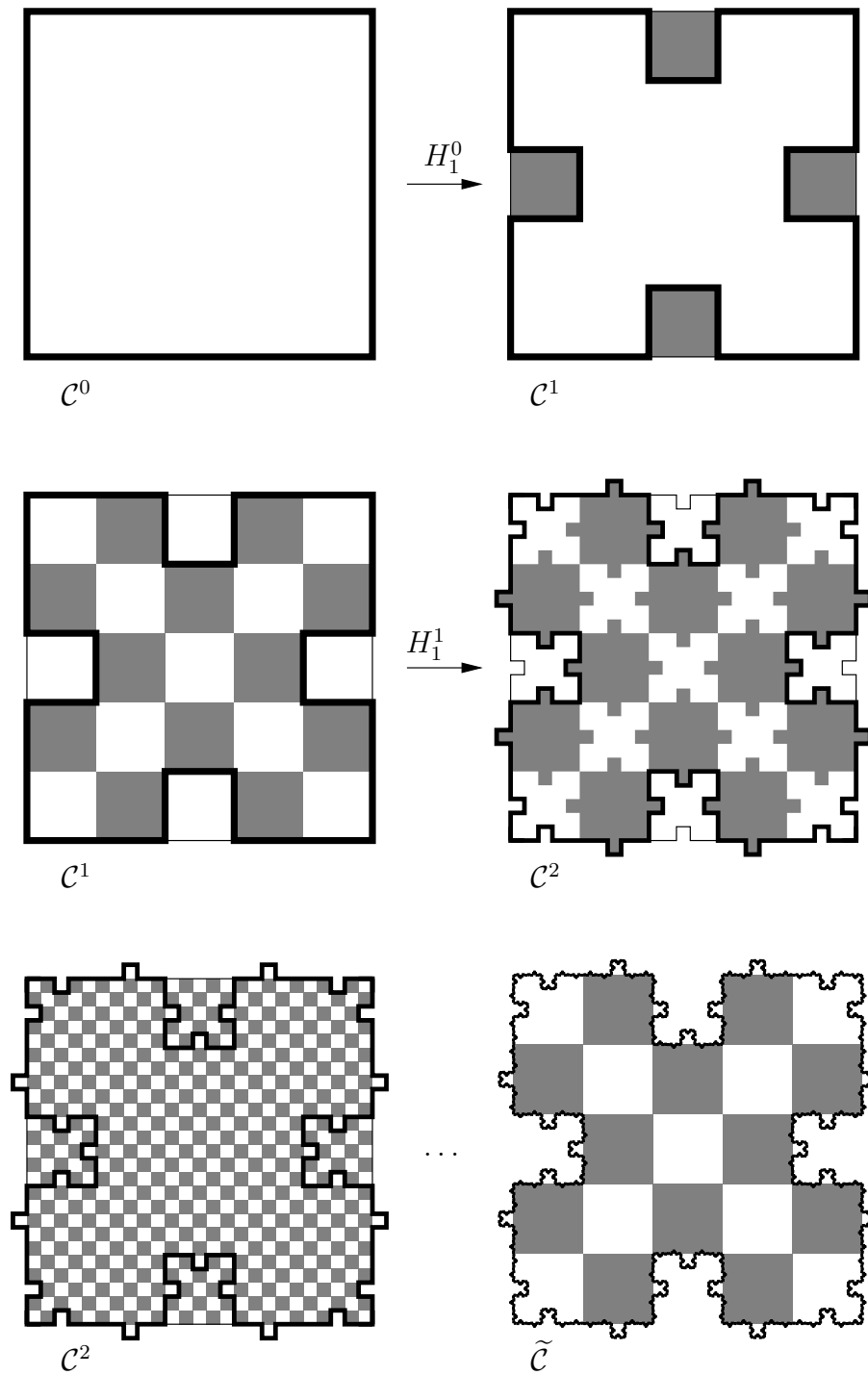


FIGURE 21. Iterative construction of an invariant curve.

It is straightforward to check that the extended real line $\mathcal{C} := \widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ is f -invariant and contains all postcritical points $0, 1, \infty, -1$ of f .

As in Figure 1 we represent the sphere as a pillow, i.e., two squares glued together along their boundary. The boundary of the pillow (i.e., the boundary along which the two squares were glued together) represents the curve \mathcal{C} , the two squares represent the 0-tiles, one of which is colored white, the other black.

The map f can then be described as follows. Each of the two sides of the pillow is divided into 5×5 squares, which are colored in a checkerboard fashion. The map g acts by mapping each small white square to the white side of the pillow, and each small black square to the black side. The two sides of the pillow are the 0-tiles with respect to \mathcal{C} ; the 4 vertices of the pillow are the postcritical points in this model. The small squares are the 1-tiles (for (f, \mathcal{C})). The coloring of the 0- and 1-tiles corresponds to a labeling map $L_{\mathbf{X}}$ as in Lemma 12.6.

There exist f -invariant Jordan curves that are isotopic to \mathcal{C} rel. $\text{post}(f)$, but distinct from \mathcal{C} . The construction of one such curve is illustrated in Figure 21. Namely, we set $\mathcal{C}^0 := \mathcal{C}$. The Jordan curve \mathcal{C}^1 is shown in the top right, as well as in the middle left picture. In the latter picture, we see that \mathcal{C}^1 consists of 1-edges, i.e., $\mathcal{C}^1 \subset f^{-1}(\mathcal{C}^0)$. Moreover, there exists an isotopy $H^0: S^2 \times I \rightarrow S^2$ rel. $\text{post}(f)$ that deforms \mathcal{C}^0 to \mathcal{C}^1 (i.e., $H_0^0 = \text{id}_{S^2}$ and $H_1^0(\mathcal{C}^0) = \mathcal{C}^1$). We also see here how the black and the white 0-tile are deformed by H_1^0 ; namely, the four small black squares in the top left of Figure 21 are part of the image of the black 0-tile (which is at the back of the pillow) under H_1^0 .

The Jordan curve $\mathcal{C}^2 := H_1^1(\mathcal{C}^1)$ consists of 2-edges, i.e., $\mathcal{C}^2 \subset g^{-2}(\mathcal{C}^0)$, see the bottom left.

The two pictures in the middle of Figure 21 indicate how H^1 deforms 1-tiles. Roughly speaking H^1 deforms each black/white 1-tile “in the same fashion” as H^0 deforms the black/white 0-tiles.

The curves \mathcal{C}^n Hausdorff converge to $\tilde{\mathcal{C}}$, which is a g -invariant Jordan curve with $\text{post}(g) \subset \tilde{\mathcal{C}}$ (see Lemma 15.12 (viii) and Proposition 15.14).

There is a conceptually different way to obtain \mathcal{C}^{n+1} from \mathcal{C}^n , which will be explained in detail in Remark 15.16. Loosely speaking, we replace each n -edge $\alpha^n \subset \mathcal{C}^n$ by $(n+1)$ -edges “in the same fashion” as the 0-edge $\alpha^0 := f^n(\alpha^0) \subset \mathcal{C}^0$ is replaced by the arc $\beta^1 \subset \mathcal{C}^1$ with the same endpoints (which are postcritical points) as α^0 . Note that $\beta^1 = H_1^0(\alpha^0)$, and that β^1 consists of 1-edges.

To prepare the proof that under suitable conditions our iteration process has an invariant Jordan curve as a Hausdorff limit, we summarize some properties of the Jordan curves \mathcal{C}^n .

Lemma 15.12. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map, and the Jordan curves \mathcal{C}^n for $n \in \mathbb{N}_0$ be defined as above. Then the following statements are true:*

- (i) $\mathcal{C}^{n+k} \subset f^{-k}(\mathcal{C}^n)$ for $n, k \in \mathbb{N}_0$.
- (ii) \mathcal{C}^{n+k} is isotopic to \mathcal{C}^n rel. $f^{-n}(\text{post}(f))$ for $n, k \in \mathbb{N}_0$.
- (iii) $\mathcal{C}^{n+k} \cap f^{-n}(\text{post}(f)) = \mathcal{C}^n \cap f^{-n}(\text{post}(f))$ for $n, k \in \mathbb{N}_0$.
- (iv) $\text{post}(f) \subset \mathcal{C}^n$ for $n \in \mathbb{N}_0$.
- (v) For $n, k \in \mathbb{N}_0$ the curve \mathcal{C}^{n+k} consists of n -edges for (f, \mathcal{C}^k) .
- (vi) For $n \in \mathbb{N}$ the curve \mathcal{C}^n is the unique Jordan curve in S^2 with $\mathcal{C}^n \subset f^{-1}(\mathcal{C}^{n-1})$ that is isotopic to \mathcal{C}^1 rel. $f^{-1}(\text{post}(f))$.
- (vii) The sequence \mathcal{C}^n , $n \in \mathbb{N}_0$, only depends on \mathcal{C}^0 and \mathcal{C}^1 and not of the choice of the initial isotopy $H = H^0$ used in the definition of the sequence.
- (viii) As $n \rightarrow \infty$ the sets \mathcal{C}^n Hausdorff converge to a closed f -invariant set $\tilde{\mathcal{C}} \subset S^2$ (i.e., $f(\tilde{\mathcal{C}}) \subset \tilde{\mathcal{C}}$) with $\text{post}(f) \subset \tilde{\mathcal{C}}$.

Proof. In the following we use the isotopies H^n as in the definition of the sequence \mathcal{C}^n , and set $h_n = H_1^n$ for $n \in \mathbb{N}_0$.

(i) It suffices to show that $\mathcal{C}^n \subset f^{-1}(\mathcal{C}^{n-1})$ for $n \in \mathbb{N}$. We prove this by induction on n ; this is clear for $n = 1$. Assume that the statement holds for some $n \in \mathbb{N}$; so $\mathcal{C}^n \subset f^{-1}(\mathcal{C}^{n-1})$. Since $h_n = H_1^n$ and $h_{n-1} = H_1^{n-1}$ are homeomorphisms with $f \circ h_n = h_{n-1} \circ f$, we have $h_n(f^{-1}(\mathcal{C}^n)) = f^{-1}(h_{n-1}(\mathcal{C}^n))$ by Lemma 10.2. Thus

$$\mathcal{C}^{n+1} = h_n(\mathcal{C}^n) \subset h_n(f^{-1}(\mathcal{C}^{n-1})) = f^{-1}(h_{n-1}(\mathcal{C}^{n-1})) = f^{-1}(\mathcal{C}^n),$$

and (i) follows.

(ii)–(iv) From the definition of H^n , the remark after the proof of Proposition 10.1, and induction on n we conclude that H^n is an isotopy rel. $f^{-n}(\text{post}(f))$. Since $H_0^n = \text{id}_{S^2}$ and

$$f^{-n}(\text{post}(f)) \subset f^{-(n+k)}(\text{post}(f))$$

for $n, k \in \mathbb{N}_0$, statements (ii) and (iii) immediately follow from this by induction on k for fixed n . Statement (iv) follows from (iii) (with $n = 0$ and $k \in \mathbb{N}_0$ arbitrary) and the fact that $\text{post}(f) \subset \mathcal{C}^0$.

(v) By (iv) we have $\#(f^{-n}(\text{post}(f)) \cap \mathcal{C}^{n+k}) \geq \# \text{post}(f) \geq 3$. In particular, the points in $f^{-n}(\text{post}(f))$ that lie on \mathcal{C}^{n+k} subdivide this curve into arcs whose endpoints lie in $f^{-n}(\text{post}(f))$ and whose interiors are

disjoint from $f^{-n}(\text{post}(f))$. Let $\alpha \subset \mathcal{C}^{n+k}$ be one of these arcs. Then we have $\text{int}(\alpha) \subset f^{-n}(\mathcal{C}^k) \setminus f^{-n}(\text{post}(f))$ by (i), and $\partial\alpha \subset f^{-n}(\text{post}(f))$. Since by Proposition 6.1 (iii) the set $f^{-n}(\mathcal{C}^k)$ is the 1-skeleton and the set $f^{-n}(\text{post}(f))$ the 0-skeleton of the cell decomposition $\mathcal{D}^n(f, \mathcal{C}^k)$, we conclude from Lemmas 4.4 and 4.5 that α is an edge in $\mathcal{D}^n(f, \mathcal{C}^k)$, i.e., an n -edge for (f, \mathcal{C}^k) . Hence \mathcal{C}^{n+k} consists of n -edges for (f, \mathcal{C}^k) .

(vi) By (i) and (ii) we know that \mathcal{C}^n for $n \in \mathbb{N}$ is a Jordan curve with $\mathcal{C}^n \subset f^{-1}(\mathcal{C}^{n-1})$ that is isotopic to \mathcal{C}^1 rel. $f^{-1}(\text{post}(f))$. Let $\widehat{\mathcal{C}} \subset f^{-1}(\mathcal{C}^{n-1})$ be another Jordan curve isotopic to \mathcal{C}^1 rel. $f^{-1}(\text{post}(f))$. Then \mathcal{C}^n and $\widehat{\mathcal{C}}$ are isotopic to each other rel. $f^{-1}(\text{post}(f))$. Note that $f^{-1}(\mathcal{C}^{n-1})$ is the 1-skeleton of the cell decomposition $\mathcal{D}^1(f, \mathcal{C}^{n-1})$ and $f^{-1}(\text{post}(f))$ its set of vertices. Hence by Lemma 10.12 we have $\widehat{\mathcal{C}} = \mathcal{C}^n$, and the uniqueness statement for \mathcal{C}^n follows.

(vii) It follows from (vi) and induction on n that \mathcal{C}^n is uniquely determined by \mathcal{C}^0 and \mathcal{C}^1 .

(viii) Pick some visual metric d for f , and let $\Lambda > 1$ be the expansion factor of d . By Lemma 10.3 the diameters of the tracks of the isotopy H^n are bounded by $C\Lambda^{-n}$, where C is a fixed constant. Since $H_0^n = \text{id}_{S^2}$ and $\mathcal{C}^{n+1} = H_1^n(\mathcal{C}^n)$ for $n \in \mathbb{N}_0$ this implies that $\text{dist}_d^N(\mathcal{C}^n, \mathcal{C}^{n+1}) \leq C\Lambda^{-n}$ for $n \in \mathbb{N}_0$. It follows that the sequence \mathcal{C}^n is a Cauchy sequence with respect to Hausdorff distance. Now the space of all non-empty closed subsets of a compact metric space is compact if it is equipped with the Hausdorff distance. Hence there exists a subsequence of the sequence \mathcal{C}^n that converges in the Hausdorff sense to a non-empty closed set $\widetilde{\mathcal{C}} \subset S^2$. Since \mathcal{C}^n is a Cauchy sequence, it follows that $\mathcal{C}^n \rightarrow \widetilde{\mathcal{C}}$ as $n \rightarrow \infty$ in the Hausdorff sense. Since $\text{post}(f) \subset \mathcal{C}^n$ for all $n \in \mathbb{N}_0$ by (iv), we have $\text{post}(f) \subset \widetilde{\mathcal{C}}$.

It remains to show that $\widetilde{\mathcal{C}}$ is f -invariant. To see this, let $p \in \widetilde{\mathcal{C}}$ be arbitrary. Then there exists a sequence (p_n) of points in S^2 such that $p_n \in \mathcal{C}^n$ for $n \in \mathbb{N}_0$ and $p_n \rightarrow p$ as $n \rightarrow \infty$. By continuity of f we have $f(p_n) \rightarrow f(p)$ as $n \rightarrow \infty$. Moreover, (i) implies that $f(p_n) \in \mathcal{C}^{n-1}$ for $n \in \mathbb{N}$. Hence $f(p) \in \widetilde{\mathcal{C}}$, and so the set $\widetilde{\mathcal{C}}$ is indeed f -invariant. \square

As an application of the preceding setup we prove a statement that gives a necessary and sufficient condition for the map \widehat{f} in Theorem 1.3 to be combinatorially expanding.

Proposition 15.13. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map and the isotopy $H^0: S^2 \times I \rightarrow S^2$ and Jordan curves \mathcal{C}^n for $n \in \mathbb{N}_0$ be defined as above.*

Then $\hat{f} = H_1^0 \circ f$ is combinatorially expanding for $\mathcal{C}^1 = \mathcal{C}'$ if and only if there exists $n \in \mathbb{N}$ such that no n -tile for (f, \mathcal{C}^0) joins opposite sides of \mathcal{C}^n .

Proof. Let H^n for $n \in \mathbb{N}_0$ be the isotopies used in the definition of the curves \mathcal{C}^n . Set $h_n := H_1^n$. Then $\hat{f} = h_0 \circ f$, $\mathcal{C}^{n+1} = h_n(\mathcal{C}^n)$, and $h_n \circ f = f \circ h_{n+1}$ for $n \in \mathbb{N}_0$. It follows by induction that for $n \in \mathbb{N}$ we have

$$\hat{f}^n = h_0 \circ f \circ \cdots \circ h_0 \circ f = h_0 \circ f^n \circ h_{n-1} \circ \cdots \circ h_1,$$

and so

$$h_0 \circ f^n = \hat{f}^n \circ h_1^{-1} \circ \cdots \circ h_{n-1}^{-1}.$$

Hence

$$f^{-n}(\mathcal{C}^0) = f^{-n}(h_0^{-1}(\mathcal{C}^1)) = (h_{n-1} \circ \cdots \circ h_1)(\hat{f}^{-n}(\mathcal{C}^1)).$$

Recall that the n -tiles for (f, \mathcal{C}^0) are the closures of the complementary components of $f^{-n}(\mathcal{C}^0)$, and the n -tiles for (\hat{f}, \mathcal{C}^1) the closures of the complementary components of $\hat{f}^{-n}(\mathcal{C}^1)$ (Proposition 6.1 (v)). So from the previous identity we conclude that the n -tiles for (f, \mathcal{C}^0) are precisely the images of the n -tiles for (\hat{f}, \mathcal{C}^1) under the homeomorphism $h_{n-1} \circ \cdots \circ h_1$. Note that this homeomorphism is isotopic to id_{S^2} rel. $\text{post}(f) = \text{post}(\hat{f})$ and maps \mathcal{C}^1 to \mathcal{C}^n . Thus no n -tile for (\hat{f}, \mathcal{C}^1) joins opposite sides of \mathcal{C}^1 if and only if no n -tile for (f, \mathcal{C}^0) joins opposite sides of \mathcal{C}^n .

Now \hat{f} is combinatorially expanding for \mathcal{C}^1 if and only if there exists $n \in \mathbb{N}$ such that no n -tile for (\hat{f}, \mathcal{C}^1) joins opposite sides of \mathcal{C}^1 . By what we have seen, this is the case if and only if there exists $n \in \mathbb{N}$ such that no n -tile for (f, \mathcal{C}^0) joins opposite sides of \mathcal{C}^n . \square

The Hausdorff limit $\tilde{\mathcal{C}}$ of the curves \mathcal{C}^n as provided by Lemma 15.12 (viii) will not be a Jordan curve in general. The following proposition shows that this is the case if the map $\hat{f} = H_1^0 \circ f$ is combinatorially expanding for \mathcal{C}^1 . Actually, one can also show that this condition is also necessary for $\tilde{\mathcal{C}}$ to be a Jordan curve, but we will not present the proof for this statement as it is somewhat involved.

Proposition 15.14 (Iterative procedure to obtain an invariant curve).

Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map and the isotopy $H^0: S^2 \times I \rightarrow S^2$ and Jordan curves \mathcal{C}^n for $n \in \mathbb{N}_0$ be defined as above.

If $\hat{f} = H_1^0 \circ f$ is combinatorially expanding for $\mathcal{C}^1 = \mathcal{C}'$, then \mathcal{C}^n Hausdorff converges to a Jordan curve $\tilde{\mathcal{C}} \subset S^2$ as $n \rightarrow \infty$. In this case,

the curve $\tilde{\mathcal{C}}$ is f -invariant and $\text{post}(f) \subset \tilde{\mathcal{C}}$. Furthermore $\tilde{\mathcal{C}}$ is isotopic to \mathcal{C}^1 rel. $f^{-1}(\text{post}(f))$.

Proof. Suppose that $\hat{f} = H_1^0 \circ f$ is combinatorially expanding for \mathcal{C}^1 . From Theorem 1.3 it follows that there exists an f -invariant Jordan curve $\tilde{\mathcal{C}} \subset S^2$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$ that is isotopic to \mathcal{C}^0 rel. $\text{post}(f)$ and isotopic to \mathcal{C}^1 rel. $f^{-1}(\text{post}(f))$. Let $K^0: S^2 \times I \rightarrow S^2$ be an isotopy rel. $\text{post}(f)$ that deforms $\tilde{\mathcal{C}}$ to \mathcal{C}^0 ; so $K_0^0 = \text{id}_{S^2}$ and $K_1^0(\tilde{\mathcal{C}}) = \mathcal{C}^0$. Using Proposition 10.1 repeatedly, we can find isotopies $K^n: S^2 \times I \rightarrow S^2$ rel. $f^{-1}(\text{post}(f))$ with $K_0^n = \text{id}_{S^2}$ such that $f \circ K_1^n = K_1^{n-1} \circ f$ for $n \in \mathbb{N}$.

Claim. $\tilde{\mathcal{C}}^n := K_1^n(\tilde{\mathcal{C}}) = \mathcal{C}^n$ for all $n \in \mathbb{N}_0$.

We prove this claim by induction on n ; it follows from the choice of K^0 for $n = 0$. Assume that the statement is true for some $n \in \mathbb{N}_0$. Then $K_1^n(\tilde{\mathcal{C}}) = \mathcal{C}^n$, and so by Lemma 10.2 we have

$$\tilde{\mathcal{C}}^{n+1} = K_1^{n+1}(\tilde{\mathcal{C}}) \subset K_1^{n+1}(f^{-1}(\tilde{\mathcal{C}})) = f^{-1}(K_1^n(\tilde{\mathcal{C}})) = f^{-1}(\mathcal{C}^n).$$

Since K^{n+1} is an isotopy rel. $f^{-1}(\text{post}(f))$, the curve $\tilde{\mathcal{C}}^{n+1}$ is isotopic to $\tilde{\mathcal{C}}$ and hence to \mathcal{C}^1 rel. $f^{-1}(\text{post}(f))$. So Lemma 15.12 (vi) implies that $\tilde{\mathcal{C}}^{n+1} = \mathcal{C}^{n+1}$ proving the claim.

By Lemma 10.3 the maps K_1^n converge uniformly to the identity on S^2 . Hence $\mathcal{C}^n = K^n(\tilde{\mathcal{C}})$ Hausdorff converges to the Jordan curve $\tilde{\mathcal{C}}$ as $n \rightarrow \infty$. The statement follows. \square

Remark 15.15. If $f: S^2 \rightarrow S^2$ is an expanding Thurston map, then every f -invariant Jordan curve $\tilde{\mathcal{C}}$ with $\text{post}(f) \subset \tilde{\mathcal{C}}$ can be obtained by our iterative procedure. Indeed, suppose that $\tilde{\mathcal{C}}$ is such a curve. Trivially, we can then take $\mathcal{C} = \mathcal{C}^0 = \tilde{\mathcal{C}}$, $\mathcal{C}' = \mathcal{C}^1 = \tilde{\mathcal{C}}$, and $H_t^0 = \text{id}_{S^2}$ for $t \in I$. Then $\mathcal{C}^n = \tilde{\mathcal{C}}$ for all $n \in \mathbb{N}_0$ and so $\mathcal{C}^n \rightarrow \tilde{\mathcal{C}}$ as $n \rightarrow \infty$.

Actually, a much stronger statement is true. Namely, we can start with *any* Jordan curve \mathcal{C} in the same isotopy class rel. $\text{post}(f)$ as $\tilde{\mathcal{C}}$. Suppose that \mathcal{C} is such a curve. First, we claim that then there exists a unique Jordan curve $\mathcal{C}' \subset f^{-1}(\mathcal{C})$ that is isotopic to $\tilde{\mathcal{C}}$ rel. $f^{-1}(\text{post}(f))$. To see this, let $K^0: S^2 \times I \rightarrow S^2$ be an isotopy rel. $\text{post}(f)$ with $K_0^0 = \text{id}_{S^2}$ and $K_1^0(\tilde{\mathcal{C}}) = \mathcal{C}$. By Proposition 10.1 we can lift K^0 by f to an isotopy K^1 rel. $f^{-1}(\text{post}(f))$ with $K_0^1 = \text{id}_{S^2}$ and $K_t^0 \circ f = f \circ K_t^1$ for all $t \in I$. Then the Jordan curve $\mathcal{C}' := K_1^1(\tilde{\mathcal{C}})$ satisfies

$$\mathcal{C}' = K_1^1(\tilde{\mathcal{C}}) \subset K_1^1(f^{-1}(\tilde{\mathcal{C}})) = f^{-1}(K_1^0(\tilde{\mathcal{C}})) = f^{-1}(\mathcal{C}).$$

Here we used $\tilde{\mathcal{C}} \subset f^{-1}(\tilde{\mathcal{C}})$ and Lemma 10.2. This shows existence of a curve \mathcal{C}' with the desired properties. Uniqueness of \mathcal{C}' follows from Lemma 10.12.

Define $H: S^2 \times I \rightarrow S^2$ by setting $H_t = K_t^1 \circ (K_t^0)^{-1}$ for $t \in I$. Then H is an isotopy rel. $\text{post}(f)$ that deforms $\mathcal{C}^0 := \mathcal{C}$ into $\mathcal{C}^1 := \mathcal{C}'$. Indeed, we have $H_0 = \text{id}_{S^2}$ and

$$H_1(\mathcal{C}^0) = K_1^1((K_1^0)^{-1}(\mathcal{C})) = K_1^1(\tilde{\mathcal{C}}) = \mathcal{C}' = \mathcal{C}^1.$$

Moreover,

$$\hat{f} := H_1 \circ f = K_1^1 \circ (K_1^0)^{-1} \circ f = K_1^1 \circ f \circ (K_1^1)^{-1}.$$

Thus it follows from Lemma 12.8 that \hat{f} is combinatorially expanding for $\mathcal{C}^1 = \mathcal{C}' = K_1^1(\tilde{\mathcal{C}})$.

Define the sequence $\{\mathcal{C}^n\}$ starting from \mathcal{C}^0 and \mathcal{C}^1 as before. From Proposition 15.14 it follows that as $n \rightarrow \infty$ the curves \mathcal{C}^n Hausdorff converge to an f -invariant Jordan curve that is isotopic to \mathcal{C}^1 , and hence isotopic to $\tilde{\mathcal{C}}$, rel. $f^{-1}(\text{post}(f))$. From Theorem 1.4 it follows that the unique such curve is $\tilde{\mathcal{C}}$. Thus $\mathcal{C}^n \rightarrow \tilde{\mathcal{C}}$ in the Hausdorff sense as $n \rightarrow \infty$.

Remark 15.16. In the inductive definition of $\mathcal{C}^{n+1} = H_1^n(\mathcal{C}^n)$ one can construct \mathcal{C}^{n+1} from \mathcal{C}^n by an *edge replacement procedure* without explicitly knowing the isotopy H^n . To explain this, suppose that $n \in \mathbb{N}$, and that \mathcal{C}^n has already been constructed (starting from given curves \mathcal{C}^0 and \mathcal{C}^1). We know by Lemma 15.12 (v) that \mathcal{C}^n consists of n -edges α^n for (f, \mathcal{C}^0) . Then \mathcal{C}^{n+1} is obtained from \mathcal{C}^n by replacing each n -edge $\alpha^n \subset \mathcal{C}^n$ by a certain arc β^{n+1} with the same endpoints as α^n .

Indeed, we can set $\beta^{n+1} := H_1^n(\alpha^n) \subset \mathcal{C}^{n+1}$. Then the union of these arcs β^{n+1} is equal to \mathcal{C}^{n+1} . Moreover, since H^n is an isotopy relative to the set $f^{-n}(\text{post}(f))$ of n -vertices, and α^n is an n -edge for (f, \mathcal{C}^0) and so has n -vertices as endpoints, the arcs α^n and β^{n+1} have the same endpoints.

Now the arc β^{n+1} is the unique arc in $f^{-n}(\mathcal{C}^1)$ that is isotopic to α^n rel. $f^{-n}(\text{post}(f))$. This property often allows one to determine β^{n+1} from α^n without knowing H^n explicitly.

To see that this characterization of β^{n+1} holds, note that by Lemma 15.12 (i) we have $\beta^{n+1} \subset \mathcal{C}^{n+1} \subset f^{-1}(\mathcal{C}^n)$. Moreover, $\beta^{n+1} = H_1^n(\alpha^n)$ is isotopic to α^n rel. $f^{-n}(\text{post}(f))$.

Suppose $\tilde{\beta}^{n+1} \subset f^{-n}(\mathcal{C}^1)$ is another arc that is isotopic to α^n rel. $f^{-n}(\text{post}(f))$. Then the arcs β^{n+1} and $\tilde{\beta}^{n+1}$ have endpoints in $f^{-n}(\text{post}(f))$, but contain no other points in this set, since this is true for α^n . This and the inclusions $\beta^{n+1}, \tilde{\beta}^{n+1} \subset f^{-n}(\mathcal{C}^1)$ imply that β^{n+1} and $\tilde{\beta}^{n+1}$ are n -edges for (f, \mathcal{C}^1) (see the argument in the proof of Lemma 15.12 (v)). Since β^{n+1} and $\tilde{\beta}^{n+1}$ are isotopic relative to the set $f^{-n}(\text{post}(f))$, which

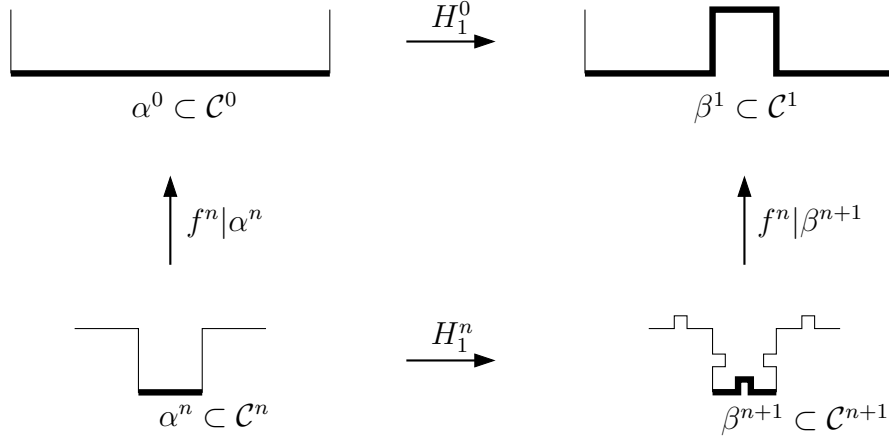


FIGURE 22. Iterative construction by replacing edges.

is the 0-skeleton of $\mathcal{D}^n(f, \mathcal{C}^1)$, it follows from the first part of the proof of Lemma 10.12 that $\beta^{n+1} = \tilde{\beta}^{n+1}$ as desired.

As we have just seen, β^{n+1} is an n -edge for (f, \mathcal{C}^1) . Since β^{n+1} has endpoints in the set $f^{-n}(\text{post}(f)) \subset f^{-(n+1)}(\text{post}(f))$ and $\beta^{n+1} \subset \mathcal{C}^{n+1} \subset f^{-(n+1)}(\mathcal{C}^0)$, a similar argument also shows that β^{n+1} consists of $(n+1)$ -edges for (f, \mathcal{C}^0) .

One can look at the arc replacement procedure $\alpha^n \rightarrow \beta^{n+1}$ from yet another point of view. Since α^n is an n -edge for (f, \mathcal{C}^0) , the map $f^n|_{\alpha^n}$ is a homeomorphism of α^n onto the 0-edge $\alpha^0 := f^n(\alpha^n) \subset \mathcal{C}^0$ for (f, \mathcal{C}^0) (Proposition 6.1 (i)). The endpoints of α^0 lie in $\text{post}(f)$. Then $\beta^1 := H_1^0(\alpha^0)$ is the unique subarc of \mathcal{C}^1 that has the same endpoints as α^0 , but contains no other points in $\text{post}(f)$ (here it is important that $\#(\mathcal{C}^1 \cap \text{post}(f)) = \#\text{post}(f) \geq 3$). Since $f^n \circ H_1^n = H_1^0 \circ f^n$ and $\beta^{n+1} = H_1^n(\alpha^n)$, the map $f^n|_{\beta^{n+1}}$ is thus a homeomorphism of β^{n+1} onto β^1 . Often, this information (together with the fact that α^n and β^{n+1} share endpoints) is enough to determine β^{n+1} uniquely. We illustrate this procedure in Figure 22. Here the map f (as well as the curves $\mathcal{C}^0, \mathcal{C}^1, \dots$ and the isotopies H^0, H^1, \dots) are as in Example 15.11, see also Figure 21.

For example, suppose that β^1 lies in *single* 0-tile X^0 for (f, \mathcal{C}^0) , i.e., in one of the Jordan regions bounded by \mathcal{C}^0 . This is not always true, but in Example 15.11 as well as the Examples 15.17 and 15.18 discussed below this is the case. Then there exists a unique n -tile X^n for (f, \mathcal{C}^0) with $\alpha^n \subset \partial X^n$ and $f^n(X^n) = X^0$; if we assign colors to tiles for (f, \mathcal{C}^0) as in Lemma 6.2, then X^n is the unique n -tile for (f, \mathcal{C}^0) that contains α^n in its boundary and has the same color as X^0 .

Consider the arc $\tilde{\beta}^{n+1} := (f^n|X^n)^{-1}(\beta^1) \subset X^n$. Then $\tilde{\beta}^{n+1}$ has the same endpoints as $(f^n|X^n)^{-1}(\alpha^0) = \alpha^n$ and is contained in $f^{-n}(\mathcal{C}^1)$. Moreover, $\tilde{\beta}^{n+1}$ is isotopic to α^n rel. $f^{-n}(\text{post}(f))$; this easily follows from Lemma 10.8, since our assumptions imply that one can find a suitable simply connected domain $\Omega \subset S^2$ that contains $\tilde{\beta}^{n+1}$ and α^n and no point in $f^{-n}(\text{post}(f))$ except the endpoints of $\tilde{\beta}^{n+1}$ and α^n . By what we have seen above, we conclude $\beta^{n+1} = \tilde{\beta}^{n+1}$, and so

$$(15.4) \quad \beta^{n+1} = (f^n|X^n)^{-1}(\beta^1).$$

In the special case under consideration, this leads to a very convenient edge replacement procedure that can be summarized as follows: Suppose the arc $\beta^1 \subset \mathcal{C}^1$ corresponding to $\alpha^0 = f^n(\alpha^n) \subset \mathcal{C}^0$ lies in a single 0-tile X^0 , and let X^n be the n -tile that contains α^n its boundary and has the same color as X^0 (so that $f^n(X^n) = X^0$). Then α^n is replaced by the arc β^{n+1} in X^n that corresponds to $\beta^1 \subset X^0$ under the homeomorphism $f^n|X^n$ of X^n onto X^0 .

The next example illustrates what happens if the map \hat{f} in Proposition 15.14 is not combinatorially expanding.

Example 15.17. Let g be the Lattès map obtained as in (1.2), where

$$\psi: \mathbb{C} \rightarrow \mathbb{C}, \quad u \mapsto \psi(u) := 3u.$$

This map was already considered in Example 13.20. See also the bottom of Figure 15. We let \mathcal{C}^0 be the boundary of the pillow. The curve $\mathcal{C}^1 \subset g^{-1}(\mathcal{C}^0)$ is drawn with a thick line in the top left of Figure 23. Clearly there is an isotopy H^0 rel. $\text{post}(g)$ (the four vertices of the pillow) that deforms \mathcal{C}^0 to \mathcal{C}^1 . Note that $\hat{g} = H_1^0 \circ g$ is not combinatorially expanding for \mathcal{C}^1 , see Figure 15. Starting with the data $\mathcal{C}^0, \mathcal{C}^1, H^0$, we inductively define Jordan curves \mathcal{C}^n as described before.

It is slightly more difficult than in Example 15.11 to see here how \mathcal{C}^{n+1} evolves from \mathcal{C}^n , since different 0-edges are replaced by different arcs (consisting of 1-edges). Namely, each 0-edge that is drawn horizontally in Figure 23 is replaced by itself. Note that every horizontal 1-edge is mapped by g to a horizontal 0-edge, thus is replaced by itself in the construction of \mathcal{C}^2 from \mathcal{C}^1 (see Remark 15.16).

Then $\mathcal{C}^n \rightarrow \tilde{\mathcal{C}}$ as $n \rightarrow \infty$ in the Hausdorff sense, where the set $\tilde{\mathcal{C}}$ is indicated on the right of Figure 23. In this case, $\tilde{\mathcal{C}}$ is not a Jordan curve and $S^2 \setminus \tilde{\mathcal{C}}$ has three components. Of course, the “self-intersections” of the limit set $\tilde{\mathcal{C}}$ can be more complicated in general.

We close this section with one more example. It gives an example of a non-trivial invariant curve that is rectifiable.

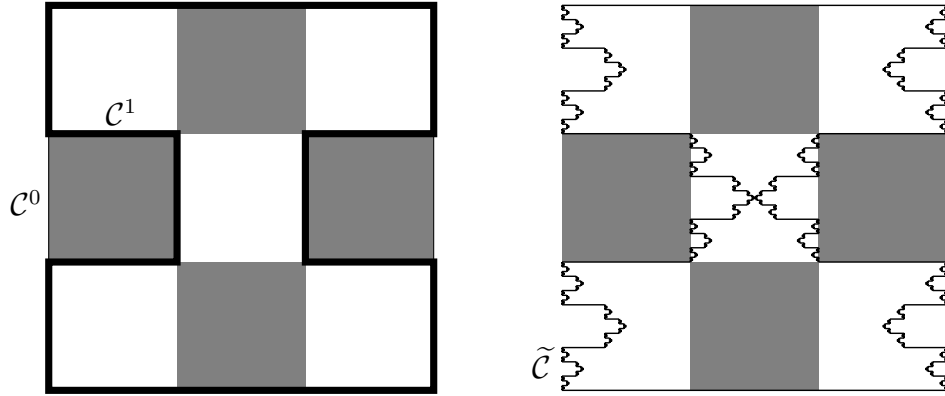


FIGURE 23. Since \hat{g} is not combinatorially expanding, $\tilde{\mathcal{C}}$ is not a Jordan curve.

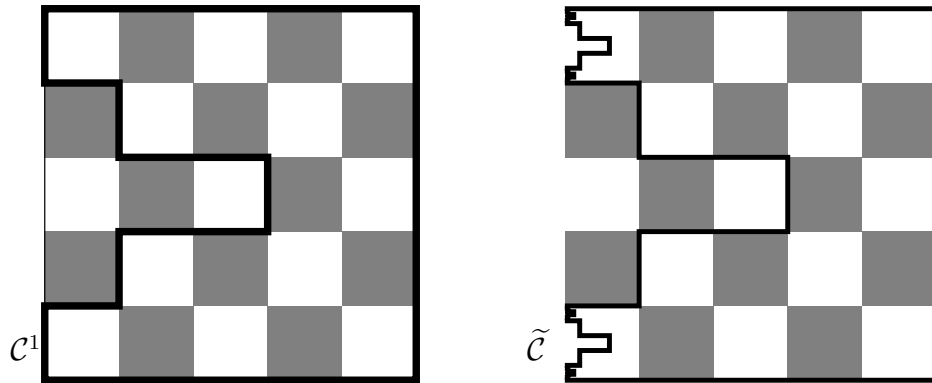


FIGURE 24. A non-trivial rectifiable invariant Jordan curve.

Example 15.18. Let f be the map from Example 15.11, i.e., the Lattès map obtained as in (1.2), where we choose $\psi: \mathbb{C} \rightarrow \mathbb{C}$, $u \mapsto \psi(u) := 5u$.

The curve $\mathcal{C} = \mathcal{C}^0$ is the boundary of the pillow as before. The curve \mathcal{C}^1 (which is isotopic rel. post(f) by an isotopy H^0) is the thick curve indicated in the left of Figure 24. Note that no 1-tile for (f, \mathcal{C}^0) connects opposite sides of \mathcal{C}^1 . Thus the sequence of curves $\{\mathcal{C}^n\}$, defined as before, Hausdorff converges to an f -invariant Jordan curve $\tilde{\mathcal{C}}$ by Proposition 15.13 and Proposition 15.14.

We briefly explain the iterative construction of the curves \mathcal{C}^n . Note that the three 0-edges on the top, bottom, and right side of the pillow are deformed by H^0 to themselves. Recall that f maps the lower left 1-tile to the white 0-tile by the map $u \mapsto 5u$ and “extends to other 1-tiles by reflection”. Note that f maps all 1-edges in \mathcal{C}^1 that are not on the left 0-edge, to one of the 0-edges on the top, right side, or bottom

of the pillow. Thus they are replaced by themselves when constructing \mathcal{C}^2 from \mathcal{C}^1 , see Remark 15.16.

The resulting f -invariant Jordan curve $\tilde{\mathcal{C}}$ is shown on the right. It is not hard to see that if the pillow is equipped with the flat metric, then the f -invariant curve $\tilde{\mathcal{C}}$ is rectifiable.

16. INVARIANT CURVES ARE QUASICIRCLES

Let (X, d_X) and (Y, d_Y) be metric spaces. A map $h: X \rightarrow Y$ is called a *quasisymmetry* or *quasisymmetric* if it is a homeomorphism and if there exists a homeomorphism $\eta: [0, \infty) \rightarrow [0, \infty)$ such that

$$\frac{d_Y(h(u), h(v))}{d_Y(h(u), h(w))} \leq \eta \left(\frac{d_Y(u, v)}{d_Y(u, w)} \right)$$

for all $u, v, w \in X$, $u \neq w$. If we want to emphasize the distortion function η here, then we call f an η -*quasisymmetry* or η -*quasisymmetric*. The metric spaces X and Y are called *quasisymmetrically equivalent* if there exists a quasisymmetry $h: X \rightarrow Y$. Two metrics d and d' on a space X are called *quasisymmetrically equivalent* if the identity map $\text{id}_X: (X, d) \rightarrow (X, d')$ is a quasisymmetry. Note that snowflake equivalence in its various incarnations (see Section 10) is stronger than quasisymmetric equivalence, since every snowflake equivalence between metric spaces is a quasisymmetry.

Suppose that S is a *metric circle*, i.e., a metric space homeomorphic to a circle. Then S is called a *quasicircle* if it is quasisymmetrically equivalent with the unit circle $\partial\mathbb{D}$ in \mathbb{C} (equipped with the Euclidean metric). A metric space X is called *doubling* if there exists $N \in \mathbb{N}$ such that every open ball of radius $0 < r \leq 2 \text{diam}(X)$ in X can be covered by at most N open balls of radius $r/2$.

According to a theorem by Tukia and Väisälä [TuV, p. 113, Thm. 4.9] a metric circle (S, d) is a quasicircle if and only if

- (i) (S, d) is *doubling*,
- (ii) S satisfies the *Ahlfors condition*: there is a constant $K \geq 1$ such that for all points $x, y \in S$, $x \neq y$, we have

$$\text{diam}_d(\gamma) \leq K d(x, y)$$

for one of the subarcs γ of S with endpoints x and y .

The *chordal metric* σ on the Riemann sphere $\hat{\mathbb{C}}$ is given by

$$\sigma(z, w) = \frac{2|z - w|}{\sqrt{1 + |z|^2} \sqrt{1 + |w|^2}}$$

for $z, w \in \mathbb{C}$ and by an appropriate limit of this expression if $z = \infty$ or $w = \infty$. If $J \subset \widehat{\mathbb{C}}$ is a Jordan curve, then, unless another metric is specified, we call J a quasicircle if J is a quasicircle if equipped with the chordal metric. Since J with the chordal metric is always a doubling metric space, a Jordan curve $J \subset \widehat{\mathbb{C}}$ is a quasicircle if and only if J satisfies the Ahlfors condition.

We are now ready to prove Theorem 1.8 stated in the introduction.

Proof of Theorem 1.8. Suppose \mathcal{C} is an f -invariant Jordan curve as in the statement, and let d be a visual metric on S^2 with expansion factor $\Lambda > 1$.

In the ensuing proof, we will consider edges for (f, \mathcal{C}) . Since \mathcal{C} is f -invariant, edges are subdivided by edges of higher order (see Proposition 11.1 (iv)). The Jordan curve \mathcal{C} is the union of all 0-edges, so this implies that \mathcal{C} is a union of n -edges for all $n \in \mathbb{N}_0$. If $n, k \in \mathbb{N}_0$ and \tilde{e} is an arbitrary $(n+k)$ -edge with $\tilde{e} \subset \mathcal{C}$, then there exists a unique n -edge e' with $\tilde{e} \subset e' \subset \mathcal{C}$. To see this pick, $p \in \text{int}(\tilde{e})$. Then there exists an n -edge e' with $p \in e' \subset \mathcal{C}$. Since e' is a union of $(n+k)$ -edges, it follows from Lemma 4.3 (ii) that $\tilde{e} \subset e'$. Uniqueness of e' is clear, because p is an interior point of each n -edge that contains \tilde{e} , and distinct n -edges have disjoint interior.

If e' is an n -edge, then the number of $(n+k)$ -edges \tilde{e} contained in e' is $\leq \# \text{post}(f) \deg(f)^k$; indeed, the images of these $(n+k)$ -edges \tilde{e} under the map f^n are distinct k -edges, and the number of k -edges is equal to $\# \text{post}(f) \deg(f)^k$ (see Lemma 4.3).

After these preliminaries, we are ready to show that \mathcal{C} equipped with (the restriction of) d is a quasicircle. We first establish that \mathcal{C} is doubling.

Let $x \in \mathcal{C}$, and $0 < r \leq 2 \text{diam}(\mathcal{C})$. In order to show that \mathcal{C} is doubling, it suffices to cover $B_d(x, r) \cap \mathcal{C}$ by a controlled number of sets of diameter $< r/4$.

It follows from Lemma 8.10 that we can find $n \in \mathbb{N}_0$ depending on r , and $k_0 \in \mathbb{N}_0$ independent of r and x with the following properties:

- (i) $r \asymp \Lambda^{-n}$,
- (ii) $\text{diam}(e) \leq r/4$ whenever e is an $(k_0 + n)$ -edge,
- (iii) $\text{dist}(e, e') \geq r$ if $n - k_0 \geq 0$ and e and e' are disjoint $(n - k_0)$ -edges.

Let E be the set of all $(n + k_0)$ -edges contained in \mathcal{C} that meet $B_d(x, r)$. Then the collection E forms a cover of $\mathcal{C} \cap B_d(x, r)$ and consists of sets of diameter $< r/4$ by (ii). Hence it suffices to find a uniform upper bound on $\#E$. If $n < k_0$, then $\#E \leq \# \text{post}(f) \deg(f)^{2k_0}$.

Otherwise, $n - k_0 \geq 0$. Then we can find an $(n - k_0)$ -edge $e \subset \mathcal{C}$ with $x \in e$. Let \tilde{e} be an arbitrary edge in E . Again we can find an $(n - k_0)$ -edge $e' \subset \mathcal{C}$ that contains \tilde{e} .

There exists a point $y \in \tilde{e} \cap B_d(x, r)$. Hence $\text{dist}(e, e') \leq d(x, y) < r$. This implies $e \cap e' \neq \emptyset$ by (iii). So whatever $\tilde{e} \in E$ is, the $(n - k_0)$ -edge $e' \subset \mathcal{C}$ meets the fixed $(n - k_0)$ -edge e . This leaves at most three possibilities for e' , namely e , and the two “neighbors” of e on \mathcal{C} . So there are three or less $(n - k_0)$ -edges that contain all the edges in E . Since each $(n - k_0)$ -edge contains at most $\# \text{post}(f) \deg(f)^{2k_0}$ edges of order $(n + k_0)$, it follows that $\#E \leq 3\# \text{post}(f) \deg(f)^{2k_0}$. In any case we get the desired bound for $\#E$.

It remains to show the Ahlfors condition. Let $x, y \in \mathcal{C}$ with $x \neq y$ be arbitrary, and let $n_0 \geq 0$ be the smallest integer for which there exist n_0 -edges $e_x \subset \mathcal{C}$ and $e_y \subset \mathcal{C}$ with $x \in e_x$, $y \in e_y$ and $e_x \cap e_y = \emptyset$. Note that n_0 is well-defined, because f is expanding and so the diameter of n -edges approaches 0 uniformly as $n \rightarrow \infty$.

Then by Lemma 8.10 (i),

$$d(x, y) \gtrsim \Lambda^{-n_0}.$$

If $n_0 = 0$, then

$$\text{diam}(\mathcal{C}) \lesssim d(x, y)$$

and there is nothing to prove. If $n_0 \geq 1$, we can find $(n_0 - 1)$ -edges $e'_x \subset \mathcal{C}$ and $e'_y \subset \mathcal{C}$ with $x \in e'_x$, $y \in e'_y$, and $e'_x \cap e'_y \neq \emptyset$. Then $e'_x \cup e'_y$ must contain one of the subarcs γ of \mathcal{C} with endpoints x and y . Hence

$$\text{diam}(\gamma) \leq \text{diam}(e'_x) + \text{diam}(e'_y) \lesssim \Lambda^{-n_0} \lesssim d(x, y).$$

Since the implicit multiplicative constants in the previous inequalities do not depend on x and y , we get a bound as desired. \square

Since the class of visual metrics for f and any of its iterates coincide (see Proposition 8.9 (v)), we may apply this theorem also to any iterate of f with an invariant Jordan curve $\mathcal{C} \supset \text{post}(f)$. In particular, the Jordan curve in Theorem 1.2 is a quasicircle if equipped with a visual metric for f .

A family of quasimetrics (possibly defined on different spaces) is called *uniformly quasimetric* if there exists a homeomorphism $\eta: [0, \infty] \rightarrow [0, \infty]$ such that each map in the family is an η -quasimetric. Obviously, each finite family of quasimetrics is uniformly quasimetric. If h is an η -quasimetric, then h^{-1} is an $\tilde{\eta}$ -quasimetric, where $\tilde{\eta}$ only depends on η ; actually, one can take $\tilde{\eta}: [0, \infty) \rightarrow [0, \infty)$ defined by $\tilde{\eta}(0) = 0$ and $\tilde{\eta}(t) = 1/\eta^{-1}(1/t)$ for $t > 0$.

This implies that if a family of maps is uniformly quasisymmetric, then the family of inverse maps is also uniformly quasisymmetric.

If X, Y, Z are metric spaces, and $h_1: X \rightarrow Y$ is η_1 -quasisymmetric and $h_2: Y \rightarrow Z$ is η_2 -quasisymmetric, then $h_2 \circ h_1$ is η -quasisymmetric, where $\eta = \eta_2 \circ \eta_1$. Hence the family of all compatible compositions of maps in two uniformly quasisymmetric families is again uniformly quasisymmetric.

An arc α equipped with some metric d is called a *quasiarc* if there exists a quasisymmetry of the unit interval $[0, 1]$ onto (α, d) . It is known that (α, d) is a quasiarc if and only if (α, d) is doubling and there exists a constant $K \geq 1$ such that $\text{diam}_d(\gamma) \leq Kd(x, y)$, whenever $x, y \in \alpha$, $x \neq y$, and γ is the subarc of α with endpoints x and y [TuV]. So quasiarcs admit a similar geometric characterization as quasicircles.

A family of arcs is said to consist of *uniform quasiarcs* if there exists a homeomorphism $\eta: [0, \infty] \rightarrow [0, \infty]$ such that for each arc α in the family there exists an η -quasisymmetry $h: [0, 1] \rightarrow \alpha$. Similarly, a family of quasicircles is said to consist of *uniform quasicircles* if there exists a homeomorphism $\eta: [0, \infty] \rightarrow [0, \infty]$ that for each quasicircle S in the family there exists an η -quasisymmetry $h: \partial\mathbb{D} \rightarrow S$. A family of quasicircles consists of uniform quasicircles if and only if the geometric conditions characterizing quasicircles, i.e., the doubling condition and the Ahlfors condition, holds with uniform parameters. A similar statement is true for families of quasiarcs [TuV].

We want to show that if the assumptions are as in Theorem 1.8, then all boundaries of tiles for (f, \mathcal{C}) are quasicircles and all edges for (f, \mathcal{C}) are quasiarcs. Actually, the family of all boundaries of tiles consists of uniform quasicircles and the family of all edges consists of uniform quasiarcs. One way to do this is to repeat the proof Theorem 1.8 and show that the geometric conditions characterizing quasiarcs and quasicircles are true for the edges and boundaries of tiles with uniform constants. We choose a different approach that is based on the following lemma which is of independent interest.

Lemma 16.1. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map, and $\mathcal{C} \subset S^2$ be an f -invariant Jordan curve with $\text{post}(f) \subset \mathcal{C}$. Suppose that S^2 is equipped with a visual metric d for f with expansion factor $\Lambda > 1$, and denote by \mathbf{X}^n for $n \in \mathbb{N}_0$ the set of n -tiles for (f, \mathcal{C}) . Then there exists a constant $C \geq 1$ with the following property:*

If $k, n \in \mathbb{N}_0$, $X^{n+k} \in \mathbf{X}^{n+k}$, and $x, y \in X^{n+k}$, then

$$(16.1) \quad \frac{1}{C}d(x, y) \leq \frac{d(f^n(x), f^n(y))}{\Lambda^n} \leq Cd(x, y).$$

In particular, the family $\mathcal{F} := \{f^n|_{X^{n+k}} : k, n \in \mathbb{N}_0, X^{n+k} \in \mathbf{X}^{n+k}\}$ is uniformly quasimetric.

Proof. In the following all cells will be for (f, \mathcal{C}) . Let $m = m_{f, \mathcal{C}}$ be as in Definition 8.5. We know by Definition 8.7 and by Lemma 8.6 (iii) that $d(x, y) \asymp \Lambda^{-m(x, y)}$, whenever $x, y \in S^2$. If $n \in \mathbb{N}_0$, then Lemma 8.6 (ii) implies that

$$m(f^n(x), f^n(y)) \geq m(x, y) - n,$$

and so

$$d(f^n(x), f^n(y)) \lesssim \Lambda^n d(x, y).$$

Here the implicit multiplicative constants are independent of x and y .

To obtain an inequality in the other direction, let $x, y \in X^{n+k} \in \mathbf{X}^{n+k}$, where $n, k \in \mathbb{N}_0$. We may assume that $x \neq y$. Then by definition of $m(x, y)$ we have $n + k \leq m(x, y) < \infty$. Let $l := m(x, y) + 1 \in \mathbb{N}$. Since $l > n + k$, the $(n + k)$ -tile X^{n+k} is subdivided by tiles of order l (Proposition 11.1 (iii)). Hence there exist l -tiles $X, Y \subset X^{n+k}$ with $x \in X$ and $y \in Y$. Then $X \cap Y = \emptyset$ by definition of $m(x, y)$. Let $X' := f^n(X)$ and $Y' := f^n(Y)$. Then by Proposition 6.1 (i) the sets X' and Y' are $(l - n)$ -tiles. Since $f^n|_{X^{n+k}}$ is injective, these tiles are disjoint, and we have $f^n(x) \in X'$ and $f^n(y) \in Y'$. So from Lemma 8.10 (i) we conclude that

$$d(f^n(x), f^n(y)) \geq \text{dist}_d(X', Y') \gtrsim \Lambda^{-(l-n)} \asymp \Lambda^n \Lambda^{-m(x, y)} \asymp \Lambda^n d(x, y).$$

Here the implicit multiplicative constants are again independent of x and y , and we get the other desired inequality.

Inequality (16.1) immediately implies that the family \mathcal{F} is uniformly quasimetric. To see this, let $k, n \in \mathbb{N}_0$ and $X^{n+k} \in \mathbf{X}^{n+k}$. Then $f^n|_{X^{n+k}}$ is a homeomorphism onto its image (see Proposition 6.1 (i)). Moreover, if $u, v, w \in X^{n+k}$, $u \neq w$, then by (16.1) we have

$$\frac{d(f^n(u), f^n(v))}{d(f^n(u), f^n(w))} \leq C^2 \frac{d(u, v)}{d(u, w)}.$$

Hence $f^n|_{X^{n+k}}$ is η -quasimetric, where $\eta(t) = C^2 t$ for $t \geq 0$. Since η is independent of the chosen map, the family \mathcal{F} is uniformly quasimetric. \square

Proposition 16.2. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map and $\mathcal{C} \subset S^2$ be an f -invariant Jordan curve with $\text{post}(f) \subset \mathcal{C}$. Suppose that S^2 is equipped with a visual metric for f , and for $n \in \mathbb{N}_0$ denote by \mathbf{X}^n the set of n -tiles and by \mathbf{E}^n the set of n -edges for (f, \mathcal{C}) .*

Then the family $\{\partial X : n \in \mathbb{N}_0 \text{ and } X \in \mathbf{X}^n\}$ consists of uniform quasicircles and the family $\{e : n \in \mathbb{N}_0 \text{ and } e \in \mathbf{E}^n\}$ of uniform quasarcs.

In particular, edges for (f, \mathcal{C}) are quasiarcs and the boundaries of all tiles are quasicircles.

Proof. By Theorem 1.8 there exists a quasimetric map $h: \partial\mathbb{D} \rightarrow \mathcal{C}$. Let X be an arbitrary tile for (f, \mathcal{C}) , say an n -tile, where $n \in \mathbb{N}_0$. Then $f^n|X$ is a homeomorphism of X onto the 0-tile $f^n(X)$ (Proposition 6.1 (i)), and so

$$f^n(\partial X) = \partial f^n(X) = \mathcal{C}.$$

By Lemma 16.1 the map $f^n|X$, and hence also the map $(f^n|X)^{-1}$, is a quasimetric. It follows that $(f^n|X)^{-1} \circ h$ is a quasimetric map from $\partial\mathbb{D}$ onto ∂X . Hence ∂X is a quasicircle. Actually, the family of these quasicircles ∂X is uniform, since the family of all relevant maps $(f^n|X)^{-1} \circ h$ is uniformly quasimetric as follows from Lemma 16.1.

The proof that the family $\{e : n \in \mathbb{N}_0 \text{ and } e \in \mathbf{E}^n\}$ consists of uniform quasiarcs runs along the same lines. First note that each 0-edge is a subarc of \mathcal{C} , and hence corresponds to a subarc of $\partial\mathbb{D}$ under the quasimetric h . Since this subarc can be mapped to the unit interval $[0, 1]$ by a bi-Lipschitz homeomorphism, each 0-edge is quasimetrically equivalent to $[0, 1]$ and hence a quasiarc.

Now let e be an arbitrary edge for (f, \mathcal{C}) , say an n -edge, where $n \in \mathbb{N}_0$. Then $f^n|e$ is a homeomorphism of e onto the 0-edge $f^n(e)$ (Proposition 6.1 (i)). Moreover, there exists an n -tile X with $e \subset X$ (Lemma 5.1). Then $f^n|e$ is the restriction of the map $f^n|X$ to e , and it follows from Lemma 16.1 that $f^n|e$ is a quasimetric. Hence e is quasimetrically equivalent to a 0-edge and hence a quasiarc.

Lemma 16.1 actually implies that the family $\{f^n|e : n \in \mathbb{N}_0 \text{ and } e \in \mathbf{E}^n\}$ is uniformly quasimetric. So each edge is quasimetrically equivalent to a 0-edge by a quasimetric in a uniformly quasimetric family. Since there are only finitely many 0-edges, this implies that the family of all edges for (f, \mathcal{C}) consists of uniform quasiarcs. \square

A metric space (X, d) is called *linearly locally connected* (often abbreviated as *LLC*) if there exists a constant $C \geq 1$ such that the following two conditions are satisfied:

- (i) If $p \in X$, $r > 0$, and $x, y \in B_d(x, r)$, $x \neq y$, then there exists a continuum $E \subset X$ with $x, y \in E$ and $E \subset B_d(p, Cr)$.
- (ii) If $p \in X$, $r > 0$, and $x, y \in X \setminus B_d(x, r)$, $x \neq y$, then there exists a continuum $E \subset X$ with $x, y \in E$ and $E \subset X \setminus B_d(p, r/C)$.

The following proposition shows that a 2-sphere equipped with a visual metric for an expanding Thurston map is linearly locally connected.

Proposition 16.3. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map, and suppose that S^2 is equipped with a visual metric d for f . Then the following statements are true:*

- (i) *There exists a constant $C' \geq 1$ such that any two points $x, y \in S^2$ can be joined by a path α in S^2 with*

$$(16.2) \quad \text{diam}(\alpha) \leq C' d(x, y).$$

- (ii) *There exists a constant $C \geq 1$ with the following property: If $p \in S^2$, $r > 0$, and $x, y \in \overline{B}(p, 2r) \setminus B(p, r)$, then there exists a path γ in S^2 joining x and y with*

$$(16.3) \quad \gamma \subset \overline{B}(p, Cr) \setminus B(p, r/C).$$

- (iii) *(S^2, d) is linearly locally connected.*

The statement (i)–(iii) are not logically independent, but one can show the implications (ii) \Rightarrow (iii) \Rightarrow (i).

Proof. Let $\Lambda > 1$ be the expansion factor of d . Then for some Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$ we have $d(u, v) \asymp \Lambda^{-m_{f,\mathcal{C}}(u,v)}$ for $u, v \in S^2$ (see Definitions 8.5 and 8.7). In the following all cells will be for (f, \mathcal{C}) .

(i) Let $x, y \in S^2$, $x \neq y$, be arbitrary, and $n = m_{f,\mathcal{C}}(x, y) \in \mathbb{N}_0$. Then there exist n -tiles X and Y with $x \in X$, $y \in Y$, and $X \cap Y \neq \emptyset$. Since X and Y are Jordan regions, we can find a path α in $X \cup Y$ that joins x and y . Then by Lemma 8.10 we have

$$\text{diam}(\alpha) \leq \text{diam}(X) + \text{diam}(Y) \lesssim \Lambda^{-n} \asymp d(x, y),$$

where the implicit multiplicative constants are independent of x and y . Statement (i) follows.

(ii) Let $p \in S^2$, $r > 0$, and $x, y \in \overline{B}(p, 2r) \setminus B(p, r)$. In the following all implicit multiplicative constant will be independent of these initial choices of p , r , x , and y .

Define

$$n := \max\{m_{f,\mathcal{C}}(p, x), m_{f,\mathcal{C}}(p, y)\} + 1.$$

Then

$$\Lambda^{-n} \asymp \min\{d(p, x), d(p, y)\} \asymp r.$$

Let X, Y, Z be n -tiles with $x \in X$, $y \in Y$, and $p \in Z$. Then by definition of n we have $X \cap Z = \emptyset$ and $Y \cap Z = \emptyset$.

Since f is expanding, we can choose $k_0 \in \mathbb{N}_0$ as in (8.4). In particular, every connected set of k_0 -tiles joining opposite sides of \mathcal{C} must contain at least 10 k_0 -tiles.

Consider set the $U^{n+k_0}(p)$ as defined in (8.11). Then $f^n(U^{n+k_0}(p))$ is connected, and consists of k_0 -tiles. This set cannot join opposite sides of \mathcal{C} ; for otherwise, we could find a connected set of six k_0 -tiles with this property (see the proof Lemma 8.13 for a similar reasoning). This is impossible by definition of k_0 . Hence $f^n(U^{n+k_0}(p))$ is contained in a 0-flower (Lemma 7.7) which implies that $U^{n+k_0}(p)$ is contained in an n -flower (Lemma 7.3 (iii)). So there exists an n -vertex v with $p \in U^{n+k_0}(p) \subset W^n(v)$. Since Z contains p , this tile must be one of the n -tiles forming the cycle of v . So $v \in Z$, and $v \notin X, Y$. This in turn implies that X and Y do not meet $W^n(v)$ (see Lemma 7.2 (iii)).

Pick a path α in S^2 that joins x and y and satisfies (16.2). By Lemma 14.4 we can find a set M of n -tiles that forms an e -chain joining X and Y so that each tile in M has non-empty intersection with α . Pick n -vertices $x' \in \partial X$, $y' \in \partial Y$. Since X and Y do not contain v , we have $x', y' \neq v$. Consider the graph $G_M = \{\partial U : U \in M\}$. It consists of n -edges, is connected, has no cut points (Lemma 14.2), and contains x' and y' as vertices. Hence there exists an edge path in G_M joining x' and y' whose underlying set β does not contain v . Then this edge path does not contain any edge in the cycle of v and so $\beta \cap W^n(v) = \emptyset$. Let γ be the path in S^2 that is obtained by running from x to x' along some path in X , then from x' to y' along β , and then from y' to y along some path in Y . Then γ joins x and y .

Since the sets X, Y, β have empty intersection with $W^n(v)$ and hence with $U^{n+k_0}(p)$, it follows that $\gamma \cap U^{n+k_0}(p) = \emptyset$. So by Lemma 8.12 we have

$$\text{dist}(p, \gamma) \gtrsim \Lambda^{-(n+k_0)} \asymp \Lambda^{-n} \asymp r.$$

Hence there exists a constant $C_1 \geq 1$ independent of the initial choices such that

$$\gamma \cap B(p, r/C_1) = \emptyset.$$

The set γ can be covered by n -tiles that meet α . Since

$$\text{diam}(\alpha) \leq C'd(x, y) \leq 4C'r \lesssim r,$$

and

$$\max\{\text{diam}(U) : U \text{ is an } n\text{-tile}\} \lesssim \Lambda^{-n} \asymp r,$$

we conclude that

$$\text{diam}(\gamma) \leq \text{diam}(\alpha) + 2 \max\{\text{diam}(U) : U \text{ is an } n\text{-tile}\} \lesssim r.$$

Since the initial point x of γ has distance $\leq 2r$ from p , it follows that there exists a constant $C_2 \geq 1$ independent of the initial choices such that $\gamma \subset B(p, C_2r)$. Setting $C = \max\{C_1, C_2\}$, get the inclusion (16.3).

(iii) To show that (S^2, d) is linearly locally connected, we verify the two relevant conditions; here we can use possibly different constants C in each of the conditions.

Let $p \in X$, $r > 0$, and $x, y \in B(p, r)$, $x \neq y$, be arbitrary. Choose a path α as in (16.2), and define $E := \alpha$ and $C = 2C' + 1$. Then $x, y \in E$, and, since $\text{diam}(\alpha) \leq C'd(x, y) \leq 2C'r$, we have

$$E \subset B(p, r + \text{diam}(\alpha)) \subset B(p, Cr).$$

The first of the *LLC*-condition follows.

For the second condition, let $p \in X$, $r > 0$, and $x, y \in X \setminus B(p, r)$ with $x \neq y$ be arbitrary. Let α be a path in S^2 joining x and y . If $\alpha \cap B(p, r) = \emptyset$, define $E := \alpha$. Then E is a continuum with $x, y \in E$ and $E \subset X \setminus B(p, r)$.

If α meets $B(p, r)$, then, as we travel from x to y along α , there exists a first point with $x' \in \overline{B}(p, 2r)$. Note that if $x \in \overline{B}(p, 2r)$, then $x' = x$, and $d(p, x') = 2r$ otherwise. In any case, $x' \in \overline{B}(p, 2r) \setminus B(p, r)$. Let α_x be the subpath of α obtained by traveling along α starting from x until we reach x' . Then $\alpha_x \subset X \setminus B(p, r)$.

Traveling along α in the opposite direction starting from y , we define a point $y' \in \overline{B}(p, 2r) \setminus B(p, r)$ and a subpath $\alpha_y \subset X \setminus B(p, r)$ of α joining y and y' similarly. Then $x', y' \in \overline{B}(p, 2r) \setminus B(p, r)$. Hence by (ii) there exists a path γ in S^2 that joins x' and y' and satisfies $\gamma \subset X \setminus B(p, r/C)$. Here $C \geq 1$ is a constant independent of the initial choices. Now define $E = \alpha_x \cup \gamma \cup \alpha_y$. Then E is a continuum with $x, y \in E$ and $E \subset X \setminus B(p, r/C)$. It follows that the second *LLC* condition is satisfied as well. \square

17. PERIODIC CRITICAL POINTS

Let $f: S^2 \rightarrow S^2$ be a branched covering map on a 2-sphere S^2 . A point $p \in S^2$ is called *periodic* if there exists $n \in \mathbb{N}$ such that $f^n(p) = p$. The smallest n for which this is true is called the *period* of the periodic point. The point p is called *pre-periodic* if there exists $n \in \mathbb{N}_0$ such that $f^n(p)$ is periodic.

The following lemma is well-known.

Lemma 17.1. *Let $f: S^2 \rightarrow S^2$ be a branched covering map. Then f has no periodic critical points if and only if there exists $N \in \mathbb{N}$ such that*

$$\deg_{f^n}(p) \leq N$$

for all $p \in S^2$ and all $n \in \mathbb{N}$.

Proof. Note that for $p \in S^2$ and $n \in \mathbb{N}$ we have

$$(17.1) \quad \deg_{f^n}(p) = \prod_{k=0}^{n-1} \deg_f(f^k(p))$$

So if p is a periodic critical point of period l , say, and $d = \deg_f(p) \geq 2$, then

$$\deg_{f^n}(p) \geq d^{\lceil n/l \rceil} \geq 2^{n/l} \rightarrow \infty$$

as $n \rightarrow \infty$. Hence $\deg_{f^n}(p)$ is not uniformly bounded.

If f has no periodic critical point, then the orbit $p, f(p), f^2(p), \dots$ of a point $p \in S^2$ can contain each critical point at most once. Hence by (17.1) we have

$$\deg_{f^n}(p) \leq N := \prod_{c \in \text{crit}(f)} \deg_f(c).$$

Note that the last product is finite, because f has only finitely many critical points. \square

Theorem 17.2. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map. Then S^2 equipped with a visual metric for f is doubling if and only if f has no periodic critical points.*

Actually, if f has no periodic critical points and d is a visual metric, then (S^2, d) is not only doubling, but even Ahlfors regular (see Proposition 20.10).

Proof. Assume first that f has no periodic critical points. By Theorem 1.2 there exists an iterate $F = f^n$ and an F -invariant Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) = \text{post}(F) \subset \mathcal{C}$. Then F is also an expanding Thurston map (Lemma 8.4) and it has no critical points as easily follows from (3.1). It suffices to show that S^2 is doubling with a visual metric for F , because the class of visual metrics for f and F agree (Proposition 8.9 (v)).

Fix such a visual metric for F , and denote by $\Lambda > 1$ its expansion factor. In the following cells will refer to cells for (F, \mathcal{C}) .

Then by Lemma 17.1, there exists $N \in \mathbb{N}$ such that $\deg_{F^k}(p) \leq N$ for all $p \in S^2$ and $k \in \mathbb{N}_0$. This implies that the closure of every k -flower consists of at most $2N$ tiles of order k (see Lemma 7.2).

To establish that S^2 is doubling we now proceed similarly as in the proof of Theorem 1.8. Let $x \in S^2$ and $0 < r \leq 2 \text{diam}(S^2)$ be arbitrary. We have to cover $B(x, r)$ by a controlled number of sets of diameter $< r/4$. Again, using Lemma 8.10, we can find $n \in \mathbb{N}_0$ depending on r , and $k_0 \in \mathbb{N}_0$ independent of r and x with the following properties:

- (i) $r \asymp \Lambda^{-n}$,
- (ii) $\text{diam}(X) \leq r/4$ whenever X is an $(k_0 + n)$ -tile,
- (iii) $\text{dist}(X, Y) \geq r$ whenever $n - k_0 \geq 0$ and X and Y are disjoint $(n - k_0)$ -tiles.

Let T be the set of all $(n + k_0)$ -tiles that meet $B(x, r)$. Then the collection T forms a cover of $B(x, r)$ and consists of sets of diameter $< r/4$ by (ii). Hence it suffices to find a uniform upper bound on $\#T$, independent of x and r . If $n < k_0$, then $\#T \leq 2 \deg(F)^{2k_0}$ (see Proposition 6.1 (iv)) and we have such a bound.

Otherwise, $n - k_0 \geq 0$. Pick an $(n - k_0)$ -tile X with $x \in X$. If Z is an arbitrary $(n + k_0)$ -tile in T , then we can find a unique $(n - k_0)$ -tile Y that contains Z (here we use that \mathcal{C} is F -invariant and so each tile is subdivided by tiles of higher order).

There exists a point $y \in Z \cap B(x, r)$. Hence $\text{dist}(X, Y) \leq d(x, y) < r$. This implies $X \cap Y \neq \emptyset$ by (iii). So whatever $Z \in T$ is, the corresponding $(n - k_0)$ -tile $Y \supset Z$ meets the fixed $(n - k_0)$ -tile X . Hence Y must share an $(n - k_0)$ -vertex v with X which implies $Y \subset \overline{W^{n-k_0}(v)}$. Since $\overline{W^{n-k_0}(v)}$ consists of at most $2N$ tiles of order $(n - k_0)$, and the number of $(n - k_0)$ -vertices in X is equal to $\#\text{post}(F)$, this leaves at most $2N\#\text{post}(f)$ possibilities for Y .

Since every $(n - k_0)$ -tile contains at most $2 \deg(f)^{2k_0}$ tiles of order $(n + k_0)$, it follows that $\#T \leq 4N\#\text{post}(f) \deg(f)^{2k_0}$. So we get a uniform bound as desired.

To show the other implication we use the following fact about doubling spaces, which is easy to show: in every ball there cannot be too many pairwise disjoint smaller balls of the same radius. More precisely, for every $\eta \in (0, 1)$ there is a number K such that every ball open ball of radius r contains at most K pairwise disjoint open balls of radius ηr .

Now suppose $f: S^2 \rightarrow S^2$ is an expanding Thurston map such that S^2 equipped with some visual metric is doubling. Pick a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$. In the following cells will be for (f, \mathcal{C}) . Let $p \in S^2$ and $n \in \mathbb{N}$. In order to show that f has no periodic critical points it suffices to give a uniform bound on $d = \deg_{f^n}(p)$ (see Lemma 17.1). For this we may assume that $\deg_{f^n}(p) \geq 2$. Then p is an n -vertex and the closure of the n -flower consists of precisely $2 \deg_{f^n}(p)$ n -tiles. These n -tiles have pairwise disjoint interior and each interior contains a ball of radius $r \asymp \Lambda^{-n}$ (see Lemma 8.13). On the other hand, $\text{diam}(\overline{W^n(p)}) \lesssim \Lambda^{-n}$. Since S^2 is doubling, it follows that the number of these tiles and hence $\deg_{f^n}(p)$ is uniformly bounded from

above by a constant independent of n and p . Hence f has no periodic critical points. \square

18. THE COMBINATORIAL EXPANSION FACTOR

In this section we will study the asymptotic rate at which the quantity D_k introduced in (7.4) grows as $k \rightarrow \infty$. Note that D_k depends on f and \mathcal{C} . If we want to emphasize this dependence, we write $D_k = D_k(f, \mathcal{C})$. We require the following lemma.

Lemma 18.1. *Let $n \in \mathbb{N}_0$, $f: S^2 \rightarrow S^2$ be a Thurston map with $\#\text{post}(f) \geq 3$, and $\mathcal{C} \subset S^2$ a Jordan curve with $\text{post}(f) \subset \mathcal{C}$. If there exists a connected set $K \subset S^2$ that joins opposite sides of \mathcal{C} and can be covered by N n -flowers for (f, \mathcal{C}) , then $D_n(f, \mathcal{C}) \leq 4N$.*

Proof. We assume first that $\#\text{post}(f) = 3$. Let K be as in the statement. By picking a point from the intersection of K with each of the three 0-edges, we can find a set $\{x, y, z\} \subset K$ such that $\{x, y, z\}$ meets opposite sides of \mathcal{C} . Since K is connected and can be covered by N n -flowers, we can find n -vertices $v_1, \dots, v_N \in S^2$ such that $x \in W^n(v_1)$, $y \in W^n(v_N)$ and $W^n(v_i) \cap W^n(v_{i+1}) \neq \emptyset$ for $i = 1, \dots, N-1$. It follows from Lemma 7.2 (i) that there exists a chain consisting of n -tiles X_1, \dots, X_{2N} joining x and y (recall the terminology discussed before Lemma 8.6). Similarly, there exists a chain X'_1, \dots, X'_{2N} of n -tiles joining x and z . The union K' of the n -tiles in these two chains is a connected set consisting of at most $4N$ n -tiles. It contains the set $\{x, y, z\}$ and hence joins opposite sides of \mathcal{C} . Thus $D_n(f, \mathcal{C}) \leq 4N$.

If $\#\text{post}(f) \geq 4$, the proof is similar and easier. In this case we can find a set $\{x, y\} \subset K$ that joins opposite sides of \mathcal{C} . By the same argument as before, we get the bound $D_n(f, \mathcal{C}) \leq 2N$. \square

Proposition 18.2. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$.*

Then the limit

$$\Lambda_0(f) = \lim_{k \rightarrow \infty} D_k(f, \mathcal{C})^{1/k}$$

exists. Moreover, this limit is independent of \mathcal{C} and we have $1 < \Lambda_0(f) < \infty$.

We call $\Lambda_0(f)$ the *combinatorial expansion factor* of f .

Proof. Set $D_k = D_k(f, \mathcal{C})$. We will first show that there exists a constant $C_1 \geq 1$ such that

$$(18.1) \quad (1/C_1)D_k \leq D_{k+1} \leq C_1 D_k$$

for all $k \in \mathbb{N}_0$. Indeed, if $k \in \mathbb{N}_0$ then there exists a connected set K joining opposite sides of \mathcal{C} that consists of D_k k -tiles (for (f, \mathcal{C})). By Lemma 7.11 we can cover K by MD_k $(k+1)$ -flowers, where $M \in \mathbb{N}$ is independent of k . Hence by Lemma 18.1 we have $D_{k+1} \leq CD_k$, where $C = 4M$. An inequality in the opposite direction follows from a similar argument again based on Lemma 7.11 and Lemma 18.1.

Now let $\tilde{\mathcal{C}} \subset S^2$ be another Jordan curve that contains $\text{post}(f)$. Set $\tilde{D}_k = D_k(f, \tilde{\mathcal{C}})$. We will show that there exists a constant $C_2 \geq 1$ such that

$$(18.2) \quad (1/C_2)D_k \leq \tilde{D}_k \leq C_2 D_k$$

for all $k \in \mathbb{N}_0$.

Let $\tilde{\delta}_0 = \delta_0(f, \tilde{\mathcal{C}}) > 0$ be defined as in (7.3) for f , $\tilde{\mathcal{C}}$, and a base metric d on S^2 . Since f is expanding, we there exists $k_0 \in \mathbb{N}_0$ such that $\text{diam}(X) < \tilde{\delta}_0/2$ whenever X is k_0 -tile for (f, \mathcal{C}) .

We can find a compact connected set \tilde{K} joining opposite sides of $\tilde{\mathcal{C}}$ that consists of \tilde{D}_k k -tiles for $(f, \tilde{\mathcal{C}})$. Then $\text{diam}(\tilde{K}) \geq \tilde{\delta}_0$ and so \tilde{K} contains two points with $d(x, y) \geq \tilde{\delta}_0$. There exist k_0 -tiles X and Y for (f, \mathcal{C}) such that $x \in X$ and $y \in Y$. By choice of k_0 we have $X \cap Y = \emptyset$, and so \tilde{K} joins k_0 -tiles for (f, \mathcal{C}) that are disjoint. Hence $f^{k_0}(\tilde{K})$ joins opposite sides of \mathcal{C} by Lemma 7.9. Every k -tile for $(f, \tilde{\mathcal{C}})$ can be covered by M k -flowers for (f, \mathcal{C}) , where M only depends on \mathcal{C} and $\tilde{\mathcal{C}}$ (Lemma 7.12). This and Lemma 7.3 imply that if $k \geq k_0$, then we can cover $f^{k_0}(\tilde{K})$ by $M\tilde{D}_k$ $(k - k_0)$ -flowers for (f, \mathcal{C}) .

So by Lemma 18.1 we have

$$D_{k-k_0} \leq 4M\tilde{D}_k,$$

and the first part of the proof implies

$$D_k \leq C_1^{k_0} D_{k-k_0} \leq 2MC_1^{k_0} \tilde{D}_k.$$

If $k \leq k_0$ we get a similar bound from the inequalities $D_k \leq 2 \deg(f)^{k_0}$ and $\tilde{D}_k \geq 1$. It follows that there exists a constant C independent of k such that

$$D_k \leq C\tilde{D}_k$$

for all $k \in \mathbb{N}_0$. An inequality in the opposite direction is obtained by reversing the roles of \mathcal{C} and $\tilde{\mathcal{C}}$ and using an inequality similar to (18.1) for \tilde{D}_k .

A consequence of (18.2) is that if the sequence $\{D_k(f, \mathcal{C})^{1/k}\}$ converges as $k \rightarrow \infty$, then $\{D_k(f, \tilde{\mathcal{C}})^{1/k}\}$ also converges and has the same limit.

So if the limit exists, it does not depend on \mathcal{C} . To show existence we may impose additional assumptions on \mathcal{C} , namely by Theorem 1.2, we may assume that \mathcal{C} is invariant for some iterate $F = f^n$ of f . Since F is also an expanding Thurston map (Lemma 8.4), it follows from Lemma 8.3 and Lemma 11.3 that the limit

$$\Lambda_0(F, \mathcal{C}) = \lim_{k \rightarrow \infty} D_k(F, \mathcal{C})^{1/k}$$

exists and that $\Lambda_0(F, \mathcal{C}) \in (1, \infty)$.

Since the k -tiles for (F, \mathcal{C}) are precisely the (nk) -tiles for (f, \mathcal{C}) we have $D_{nk} = D_k(F, \mathcal{C})$ for all $k \in \mathbb{N}_0$, and so

$$D_{nk}^{1/(nk)} = D_k(F, \mathcal{C})^{1/(nk)} \rightarrow \Lambda_0(f) := \Lambda_0(F, \mathcal{C})^{1/n} \in (1, \infty)$$

as $k \rightarrow \infty$. Invoking (18.1) it follows that $D_k^{1/k} \rightarrow \Lambda_0(f)$ as $k \rightarrow \infty$. The proof is complete. \square

If $F = f^n$ is an iterate of f , then, as was pointed out in the previous proof, we have

$$D_k(F, \mathcal{C}) = D_{nk}(f, \mathcal{C})$$

whenever $k \in \mathbb{N}_0$ and \mathcal{C} is a Jordan curve with $\text{post}(f) \subset \mathcal{C}$. This implies

$$(18.3) \quad \Lambda_0(F) = \Lambda_0(f^n) = \Lambda_0(f)^n.$$

The combinatorial expansion factor is invariant under topological conjugacy as the next statement shows.

Proposition 18.3. *Suppose $f: S^2 \rightarrow S^2$ and $g: \widehat{S}^2 \rightarrow \widehat{S}^2$ are expanding Thurston map that are topologically conjugate. Then $\Lambda_0(f) = \Lambda_0(g)$.*

Proof. By assumption there exists a homeomorphism $h: S^2 \rightarrow \widehat{S}^2$ such that $h \circ f = g \circ h$. Pick a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$ and let $\widehat{\mathcal{C}} = h(\mathcal{C})$. Then $\widehat{\mathcal{C}}$ is a Jordan curve that contains $\text{post}(g)$, and, as in the proof of Corollary 10.6, we have

$$\mathcal{D}^k(g, \widehat{\mathcal{C}}) = \{h(c) : c \in \mathcal{D}^k(f, \mathcal{C})\}$$

for $k \in \mathbb{N}_0$. This implies that $D_k(f, \mathcal{C}) = D_k(g, \widehat{\mathcal{C}})$ for all $k \in \mathbb{N}_0$ and so

$$\Lambda_0(g) = \lim_{k \rightarrow \infty} D_k(g, \widehat{\mathcal{C}})^{1/k} = \lim_{k \rightarrow \infty} D_k(f, \mathcal{C})^{1/k} = \Lambda_0(f)$$

as desired. \square

For the proof of Theorem 1.7 we need the following lemma.

Lemma 18.4. *Let $k, n \in \mathbb{N}_0$, $f: S^2 \rightarrow S^2$ be a Thurston map, and $\mathcal{C} \subset S^2$ be a Jordan curve with $\text{post}(f) \subset \mathcal{C}$.*

- (i) If $Z \subset S^2$ is a Jordan region such that $f^n|_Z$ is a homeomorphism and $f(Z)$ is a k -tile, then Z is an $(n+k)$ -tile.
- (ii) If X' is a k -tile and $p \in S^2$ is a point with $f^n(p) \in \text{int}(X')$, then there exists a unique $(n+k)$ -tile with $p \in X$ and $f^n(X) = X'$.

Proof. It is understood that tiles are for (f, \mathcal{C}) . Note that f is cellular for $(\mathcal{D}^{n+k}(f, \mathcal{C}), \mathcal{D}^k(f, \mathcal{C}))$, and the set $f^{-k}(\text{post}(f)) \supset \text{post}(f)$ of vertices of $\mathcal{D}^k(f, \mathcal{C})$ contains the set $f^n(\text{crit}(f^n)) \subset \text{post}(f)$ (the last inclusion follows from (3.1)). Hence we are in the situation of Lemma 5.4 with $\mathcal{D} = \mathcal{D}^k(f, \mathcal{C})$ and $\mathcal{D}' = \mathcal{D}^{n+k}(f, \mathcal{C})$.

Then (i) follows from the uniqueness statement of Lemma 5.4 and the definition of \mathcal{D}' in the first paragraph of the proof of this lemma.

Moreover, under the assumptions of (ii) it follows from Claim 1 in the proof of Lemma 5.4 that there exists a unique $(n+k)$ -tile X with $p \in X$. Then $f^n(X)$ is a k -tile containing $f^n(p) \in \text{int}(X')$, and so $f^n(X) = X'$. \square

Proof of Theorem 1.7. Fix a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$, and let $D_k = D_k(f, \mathcal{C})$ for $k \in \mathbb{N}_0$. In the following cells will be for (f, \mathcal{C}) .

- (i) The proof of the first part is easy. Suppose d is a visual metric with expansion factor Λ . Then there exists a constant $C \geq 1$ such that

$$\text{diam}(X) \leq C\Lambda^{-k}$$

for all k -tiles (Lemma 8.10 (ii)). Let $\delta_0 = \delta_0(f, \mathcal{C}) > 0$ be defined as in (7.3) for f, \mathcal{C} , and the metric d .

For each $k \in \mathbb{N}_0$ there exists a connected set joining opposite sides of \mathcal{C} that consists of D_k k -tiles. Hence

$$\delta_0 \leq \text{diam}(K) \leq CD_k\Lambda^{-k}.$$

Taking the k th roots here and letting $k \rightarrow \infty$, it follows that $\Lambda \leq \Lambda_0(f)$ as desired.

We break up the proof of (ii) in two parts (iia) and (iib).

- (iia) We make the additional assumption that the Jordan curve \mathcal{C} is f -invariant and $\Lambda > 1$ satisfies

$$(18.4) \quad \Lambda \leq D_1.$$

In this case, we now proceed to construct a visual metric d with expansion factor Λ that satisfies (1.1). We first introduce some terminology.

A *tile chain* P is a finite sequence of tiles X_1, \dots, X_N , where $X_j \cap X_{j+1} \neq \emptyset$ for $j = 1, \dots, N-1$. Here we do not require the tiles to be of

the same order. A *subchain* of P is a tile chain of the form X_{j_1}, \dots, X_{j_s} , where $1 \leq j_1 < \dots < j_s \leq N$. The tile chain *joins* the sets $A, B \subset S^2$ if $A \cap X_1 \neq \emptyset$ and $B \cap X_N \neq \emptyset$. It joins the points $x, y \in S^2$ if it joins $\{x\}$ and $\{y\}$. Every chain joining two sets A and B contains a *simple* subchain joining A and B , i.e., a chain that does not contain a proper subchain joining the sets. A chain X_1, \dots, X_N joining two disjoint sets A and B is simple if and only if $X_i \cap X_j = \emptyset$ whenever $0 \leq i, j \leq N+1$ and $|i - j| \geq 2$, where $X_0 = A$ and $X_{N+1} = B$.

Define the *weight* of a k -tile X^k to be

$$(18.5) \quad w(X^k) := \Lambda^{-k},$$

and the *w-length* of a tile chain P consisting of the tiles X_1, \dots, X_N as

$$\text{length}_w(P) := \sum_{j=1}^N w(X_j).$$

Now for $x, y \in S^2$ we define

$$(18.6) \quad d(x, y) := \inf_P \text{length}_w(P),$$

where the infimum is taken over all tile chains P joining x and y . Obviously, such tile chains exist and the infimum can be taken over simple tile chains P .

Claim 1. The distance function d is a visual metric with expansion factor Λ .

Symmetry and the triangle inequality immediately follow from the definition of d . Since f is expanding, we also have $d(x, x) = 0$ for $x \in S^2$.

Let $x, y \in S^2$ with $x \neq y$ be arbitrary, and define $m = m(x, y) = m_{f, \mathcal{C}}(x, y)$ (see Definition 8.5). Then there exist m -tiles X and Y with $x \in X$, $y \in Y$ and $X \cap Y \neq \emptyset$. So X, Y is a tile chain joining x and y , and thus

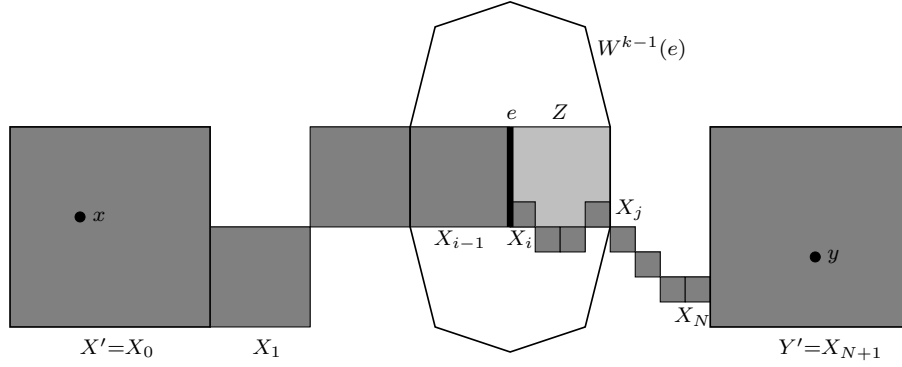
$$d(x, y) \leq w(X) + w(Y) = 2\Lambda^{-m}.$$

To prove the claim, it remains to establish a lower bound $d(x, y) \geq (1/C)\Lambda^{-m}$ for a suitable constant C independent of x and y .

Pick $(m+1)$ -tiles X' and Y' with $x \in X'$ and $y \in Y'$. Then $X' \cap Y' = \emptyset$ by definition of m . Every tile chain joining x and y contains a simple tile chain joining X' and Y' .

So let P be a simple tile chain joining X' and Y' , and suppose it consists of the tiles X_1, \dots, X_N . Let $k \in \mathbb{N}_0$ be the largest order of any tile in P . If $k \leq m+1$, then we get the favorable estimate

$$(18.7) \quad \text{length}_w(P) \geq \Lambda^{-k} \geq \Lambda^{-m-1}.$$

FIGURE 25. Replacing k -tiles by $(k - 1)$ -tiles.

Otherwise, $k > m + 1$. We want to show that then we can replace the k -tiles in P by $(k - 1)$ -tiles without increasing the w -length of the tile chain. The construction is illustrated in Figure 25.

To see this, set $X_0 = X'$, $X_{N+1} = Y'$, and let X_i , where $1 \leq i \leq N$, be the first k -tile in P . Since P is a simple chain joining X' and Y' , the tile X_i is not contained in X_{i-1} and so it has to meet ∂X_{i-1} . Since the order of X_{i-1} is $< k$, we can find a $(k - 1)$ -edge $e \subset \partial X_{i-1}$ with $e \cap X_i \neq \emptyset$. Here and below we use the fact that \mathcal{C} is f -invariant, and so cells of any order are subdivided by cells of higher order. Every $(k - 1)$ -tile meets e or is contained in the complement of the edge flower $W^{k-1}(e)$ (see Lemma 7.5 (iii)). Since tiles of order $\leq k - 1$ are subdivided into tiles of order $k - 1$, this implies that also every tile of order $\leq k - 1$ meets e or is contained in the complement of $W^{k-1}(e)$.

Now P is simple and so no tile in the “tail” $X_{i+1}, \dots, X_N, X_{N+1}$ meets e . Let $j' \in \mathbb{N}$ be the largest number such that $i \leq j' \leq N$ and all tiles $X_i, \dots, X_{j'}$ are k -tiles. Then $X_{j'+1}$ has order $\leq k - 1$. Since this tile does not meet e , it is contained in $S^2 \setminus W^{k-1}(e)$, and so the tiles $X_i, \dots, X_{j'}$ form a chain of k -tiles joining e and $S^2 \setminus W^{k-1}(e)$. Let $j \in \mathbb{N}$ be the smallest number with $i \leq j \leq j'$ such that X_j meets the complement of $W^{k-1}(e)$. Then X_i, \dots, X_j is a chain P^k of k -tiles joining e and $S^2 \setminus W^{k-1}(e)$. In particular, P^k joins two disjoint $(k - 1)$ -cells as follows from the definition of an edge flower (see Definition 7.4). Moreover, X_j is the only tile in the chain P^k that meets the complement of $W^{k-1}(e)$.

Since P^k joins disjoint $(k - 1)$ -cells, it follows from Lemma 7.10 that P^k has at least D_1 elements, and so by (18.4),

$$\text{length}_w(P^k) \geq D_1 \Lambda^{-k} \geq \Lambda^{-k+1}.$$

Let Z be the unique $(k-1)$ -tile with $Z \supset X_j$. Then $Z \cap X_{j+1} \neq \emptyset$. We also have $Z \cap e \neq \emptyset$; for otherwise $X_j \subset Z \subset S^2 \setminus W^{k-1}(e)$. Then $j > i$ and X_{j-1} meets X_j and so the complement of $W^{k-1}(e)$ contradicting the definition of j . So $Z \cap X_{i-1} \supset Z \cap e \neq \emptyset$. Thus we can replace the subchain P^k of P by the single $(k-1)$ -tile Z to obtain a chain P' joining X' and Y' . It satisfies

$$\text{length}_w(P') = \text{length}_w(P) - \text{length}_w(P^k) + w(Z) \leq \text{length}_w(P).$$

By passing to a subchain of P' we can find a simple chain P'' joining X' and Y' that contains fewer k -tiles than P and satisfies $\text{length}_w(P'') \leq \text{length}_w(P)$.

Continuing this process we can remove all k -tiles from the chain tile joining X' and Y' without increasing its w -length. If $k-1 > m+1$, we can repeat the process and remove the $(k-1)$ -tiles without increasing the w -length, etc. In the end we obtain a tile chain \tilde{P} joining X' and Y' that contains no tiles of order $> m+1$ and satisfies $\text{length}_w(\tilde{P}) \leq \text{length}_w(P)$. Thus

$$\text{length}_w(P) \geq \text{length}_w(\tilde{P}) \geq \Lambda^{-m-1}.$$

This together with the previous estimate (18.7) implies

$$d(x, y) \geq (1/\Lambda)\Lambda^{-m}.$$

This is an inequality as desired, and so d is indeed a visual metric with expansion factor Λ . In particular, this implies that d induces the given topology on S^2 (see Proposition 8.9 (ii)).

Claim 2. The visual metric d as defined in (18.6) has the expansion property (1.1).

We first show that

$$(18.8) \quad d(f(x), f(y)) \leq \Lambda d(x, y),$$

for all x, y with $d(x, y) < 1$.

Indeed, suppose $x, y \in S^2$ with $d(x, y) < 1$ are arbitrary. Let P be an arbitrary tile chain that joins x and y and consists of the tiles X_1, \dots, X_N . Assume in addition that P satisfies $\text{length}_w(P) < 1$. Then P does not contain 0-tiles and hence $f(X_1), \dots, f(X_N)$ is a tile chain joining $f(x)$ and $f(y)$. Calling the latter chain $f(P)$, we have

$$\text{length}_w(f(P)) = \Lambda \text{length}_w(P).$$

Taking the infimum over all such chains P leads to the desired inequality (18.8).

For an inequality in the other direction we now consider two cases for $x \in S^2$.

Case 1. $x \notin \text{crit}(f)$.

Then we can find an open neighborhood U of x on which f is a homeomorphism. Then $U' = f(U)$ is an open set containing $f(x)$. We can choose $\epsilon > 0$ and $\delta > 0$ such that $B_d(x, \delta) \subset U$, $B_d(f(x), \epsilon) \subset U'$, and $f(B_d(x, \delta)) \subset B_d(f(x), \epsilon)$.

Define $U_x = B_d(x, \delta)$, and let $y \in U_x$ be arbitrary. Then $d(f(x), f(y)) < \epsilon$. Consider a tile chain P' joining $f(x)$ and $f(y)$ whose w -length is close enough to $d(f(x), f(y))$ so that $\text{length}_w(P') < \epsilon$. By definition of the metric d , for every point z that lies on a tile in P' , we have $d(f(x), z) \leq \text{length}_w(P') < \epsilon$. Hence P' lies in $B_d(f(x), \epsilon) \subset U'$.

It follows that $(f|U)^{-1}$ is defined on every tile X' in P' , and so by Lemma 18.4 (i) the Jordan region $X = (f|U)^{-1}(X')$ is a tile contained in U . If k is the order of X' , then $k+1$ is the order of X . By considering these images of tiles in P' under $(f|U)^{-1}$, we get a tile chain P joining x and y with $\text{length}_w(P) = (1/\Lambda) \text{length}_w(P')$. Taking the infimum over P' we obtain

$$(18.9) \quad d(x, y) \leq (1/\Lambda) d(f(x), f(y)),$$

as desired.

Case 2. $x \in \text{crit}(f)$.

Then $x \in f^{-1}(\text{post}(f))$, and so x is a 1-vertex. Consider the flower $U = W^1(x)$, and its image $U' = f(W^1(x)) = W^0(f(x))$. These are open neighborhoods of x and $f(x)$, respectively, and the map $f|U \setminus \{x\}$ is a (non-branched) covering map of $U \setminus \{x\}$ onto $U' \setminus \{f(x)\}$ (all this follows from the considerations in the proof of Lemma 5.2). Again we can choose $\epsilon > 0$ and $\delta > 0$ such that $B_d(x, \delta) \subset U$, $B_d(f(x), \epsilon) \subset U'$, and $f(B_d(x, \delta)) \subset B_d(f(x), \epsilon)$.

Define $U_x = B_d(x, \delta)$, and let $y \in U_x$ be arbitrary. In order to show (18.9) we may assume $x \neq y$. Then $d(f(x), f(y)) < \epsilon$ and $f(x) \neq f(y)$. Consider a tile chain P' joining $f(x)$ and $f(y)$ consisting of tiles X'_1, \dots, X'_N . We can make the further assumptions that X'_1 is the only tile in this chain that contains $f(x)$ and that $\text{length}_w(P')$ is close enough to $d(f(x), f(y))$ such that $\text{length}_w(P') < \epsilon$. As before this implies that P' lies in U' . We now choose a path $\gamma: [0, N] \rightarrow U'$ with the following properties:

- (1) $\gamma(0) = f(x)$, $\gamma(N) = f(y)$ and $\gamma(t) \neq f(x)$ for $t \neq 0$,
- (2) $\gamma([i-1, i]) \subset X'_i$ for $i = 1, \dots, N$,
- (3) $\gamma(i-1/2) \in \text{int}(X'_i)$ for $i = 1, \dots, N$.

Since the tiles X'_i are Jordan regions, such a path γ can easily be obtained by first running from $f(x)$ in X'_1 to an interior point of X'_1 , then to a point in $X'_1 \cap X'_2$, then to an interior point of X'_2 , etc., and

finally to $f(y) \neq f(x)$ in X_N . Since X'_1 is the only tile in P' containing $f(x)$, this can be done so that the path never meets $f(x)$ except in its initial point.

There exists a lift α of this path (under f) with endpoints x and y , i.e., a path $\alpha: [0, N] \rightarrow U$ with $\alpha(0) = x$, $\alpha(N) = y$, and $f \circ \alpha = \gamma$. To obtain α , lift $\gamma|_{(0, N]}$ under the covering map $f|_{U \setminus \{x\}}$ such that the lift ends at y , and note that the lift has a unique continuous extension to $[0, N]$ by choosing x to be its initial point.

Using this lift α , we can construct a lift of our tile chain P' as follows. Consider a tile X'_i in P' and let k_i be its order. Put $p_i := \alpha(i - 1/2)$ and $p'_i := \gamma(i - 1/2)$. Then $f(p_i) = p'_i \in \text{int}(X'_i)$. By Lemma 18.4 (ii) there exists a unique $(k_i + 1)$ -tile X_i with $p_i \in X_i$ and $f(X_i) = X'_i$.

Note that

$$\gamma((0, N]) \subset U' \setminus \{f(x)\} = W^0(f(x)) \setminus \{f(x)\} \subset S^2 \setminus \text{post}(f)$$

and that the map

$$f: S^2 \setminus f^{-1}(\text{post}(f)) \rightarrow S^2 \setminus \text{post}(f)$$

is a covering map. This implies that $\alpha|[i - 1, i]$ is the unique lift of $\gamma|[i - 1, i]$ with $\alpha(i - 1/2) = p_i$. On the other hand, the path $\beta_i = (f|_{X_i})^{-1} \circ (\gamma|[i - 1, i])$ is also a lift of $\gamma|[i - 1, i]$ under f with $\beta_i(i - 1/2) = p_i$ by definition of X_i . Hence $\beta_i = \alpha|[i - 1, i]$ and so $\alpha([i - 1, i]) \subset X_i$. It follows that $x = \alpha(0) \in X_1$, $y = \alpha(N) \in X_N$, and $X_i \cap X_{i+1} \supset \{\alpha(i)\} \neq \emptyset$ for $i = 1, \dots, N - 1$.

Therefore, the tiles X_1, \dots, X_N form a tile chain P joining x and y . Since its tiles have an order by 1 larger the order of the corresponding tiles in P' , we have $\text{length}_w(P) = (1/\Lambda) \text{length}_w(P')$. Taking the infimum over P' we again obtain inequality (18.9).

Combining (18.8) and (18.9) we see that every point $x \in S^2$ has a neighborhood U_x such that (1.1) holds.

This concludes the proof for the existence of the visual metric d under the additional assumption in case (iia).

(iib) We now consider the general case.

We can choose an iterate of $F = f^n$ of f such that F has an F -invariant Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) = \text{post}(F) \subset \mathcal{C}$ (Theorem 1.2). Note that $D_k(f, \mathcal{C})^{1/k} \rightarrow \Lambda_0(f) > \Lambda$. Hence if n is sufficiently large, what we may assume by passing to an iterate of F , we also have

$$D_1(F, \mathcal{C}) = D_n(f, \mathcal{C}) \geq \Lambda^n.$$

This means that F is an expanding Thurston map satisfying the assumptions in (iia). Thus there exists a visual metric for F with expansion factor Λ^n that satisfies the corresponding version of (1.1). We

call this metric ρ in order to distinguish it from that metric d that we are trying to construct. Then ρ is a visual metric for F with expansion factor Λ^n , and for each $x \in S^2$ there exists an open neighborhood U_x of x such that

$$(18.10) \quad \rho(F(x), F(y)) = \rho(f^n(x), f^n(y)) = \Lambda^n \rho(x, y)$$

for all $y \in U_x$.

We now define d as

$$(18.11) \quad d(x, y) = \frac{1}{n} \sum_{i=0}^{n-1} \Lambda^{-i} \rho(f^i(x), f^i(y))$$

for $x, y \in S^2$.

Then (1.1) follows from the corresponding property (18.10) with the same sets U_x , $x \in S^2$; indeed, if $x \in S^2$ and $y \in U_x$ then by (18.10) we have

$$\begin{aligned} d(f(x), f(y)) &= \frac{1}{n} \sum_{i=0}^{n-1} \Lambda^{-i} \rho(f^{i+1}(x), f^{i+1}(y)) \\ &= \frac{1}{n} \left(\Lambda \sum_{i=0}^{n-2} \Lambda^{-(i+1)} \rho(f^{i+1}(x), f^{i+1}(y)) + \Lambda \rho(x, y) \right) \\ &= \Lambda \frac{1}{n} \sum_{i=0}^{n-1} \Lambda^{-i} \rho(f^i(x), f^i(y)) = \Lambda d(x, y). \end{aligned}$$

It remains to show that d is a visual metric for f with expansion factor Λ . It is clear that d is a metric on S^2 . Let $m = m_{f, \mathcal{C}}$ and $m_F = m_{F, \mathcal{C}}$ be as in Definition 8.5. Since ρ is a visual metric for F with expansion factor Λ^n we have

$$\rho(x, y) \asymp \Lambda^{-nm_F(x, y)} \asymp \Lambda^{-m(x, y)}$$

for all $x, y \in S^2$ by Lemma 8.6 (iv). Hence

$$d(x, y) \geq \frac{1}{n} \rho(x, y) \gtrsim \Lambda^{-m(x, y)}.$$

Moreover, by Lemma 8.6 (ii) we have

$$m(f^i(x), f^i(y)) \geq m(x, y) - i$$

and so

$$\rho(f^i(x), f^i(y)) \asymp \Lambda^{-m(f^i(x), f^i(y))} \leq \Lambda^i \Lambda^{-m(x, y)}$$

for all $i \in \mathbb{N}_0$. Hence

$$d(x, y) \lesssim \frac{1}{n} \sum_{i=0}^{n-1} \Lambda^{-m(x, y)} = \Lambda^{-m(x, y)}.$$

It follows that $d(x, y) \asymp \Lambda^{-m(x, y)}$ for all $x, y \in S^2$ where $C(\asymp)$ is independent of x and y . This shows that d is a visual metric for f with expansion factor Λ . \square

If $f: S^2 \rightarrow S^2$ is an expanding Thurston map and Λ is the expansion factor of a visual metric for f , then by Theorem 1.7 (i) we have $1 < \Lambda \leq \Lambda_0(f)$. On the other hand, statement (ii) in this theorem only guarantees the existence of a visual metric with expansion factor Λ for $1 < \Lambda < \Lambda_0(f)$. As the following example shows, this statement is optimal, since a visual metric with expansion factor $\Lambda = \Lambda_0(f)$ does not exist in general.

Example 18.5. In the following we will leave the verification of some details to the reader. The setup is very similar to Example 15.4. We use real notation and consider the map $\wp: \mathbb{R}^2 \cong \mathbb{C} \rightarrow \widehat{\mathbb{C}}$ as defined in Section 1.2. For $u_1, u_2 \in \mathbb{R}^2$ we have $\wp(u_1) = \wp(u_2)$ if and only if $u_2 = \pm u_1 + \gamma$ for $\gamma \in L = 2\mathbb{Z}^2$. The critical points of \wp are precisely the points in \mathbb{Z}^2 . Note that $\wp(\mathbb{Z}^2) = \{0, 1, \infty, -1\}$ and $\wp(\frac{1}{2}\mathbb{Z}^2 \setminus \mathbb{Z}^2) = \{\sqrt{2} - 1, 1 - \sqrt{2}, i, -i\}$. Now let

$$A = \begin{pmatrix} 2 & 2 \\ 0 & 2 \end{pmatrix}.$$

Then for $n \in \mathbb{N}$ we have

$$(18.12) \quad A^n = \begin{pmatrix} 2^n & n2^n \\ 0 & 2^n \end{pmatrix} \text{ and } A^{-n} = \begin{pmatrix} 2^{-n} & -n2^{-n} \\ 0 & 2^{-n} \end{pmatrix}.$$

Denote by $\psi: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ the map induced by left-multiplication by A , i.e., $\psi(u) = Au$ for $u \in \mathbb{R}^2$, where $u \in \mathbb{R}^2$ is considered as a column vector. Then there exists a well-defined and unique map $f: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ such that the diagram

$$\begin{array}{ccc} \mathbb{R}^2 & \xrightarrow{\psi} & \mathbb{R}^2 \\ \wp \downarrow & & \downarrow \wp \\ \widehat{\mathbb{C}} & \xrightarrow{f} & \widehat{\mathbb{C}} \end{array}$$

commutes. The map f is a branched covering map of $\widehat{\mathbb{C}}$ with

$$\text{crit}(f) = \wp(\psi^{-1}(\mathbb{Z}^2) \setminus \mathbb{Z}^2) = \wp(\tfrac{1}{2}\mathbb{Z}^2 \setminus \mathbb{Z}^2) = \{\sqrt{2} - 1, 1 - \sqrt{2}, i, -i\},$$

and

$$\text{post}(f) = \wp(\mathbb{Z}^2) = \{0, 1, \infty, -1\}.$$

Hence f is a Thurston map. Let $Q = [0, 1]^2$ be the unit square and $G \subset \mathbb{R}^2$ by the standard grid (see Example 15.4). Then $\mathcal{C} := \wp(\partial Q) = \wp(G)$ is a Jordan curve in $\widehat{\mathbb{C}}$ with $\text{post}(f) = \wp(\mathbb{Z}^2) \subset \mathcal{C}$.

Let $n \in \mathbb{N}$ and $G_n = \psi^{-n}(G)$. Then the n -tiles for (f, \mathcal{C}) are precisely the closures of the images of the complementary regions of G_n in \mathbb{R}^2 under the map \wp , i.e., precisely the sets given by

$$X^n = \wp(\psi^{-n}(\alpha + Q)),$$

where $\alpha \in \mathbb{Z}^2$. Indeed, suppose a set $X^n \subset \widehat{\mathbb{C}}$ has this form. Then X^n is a Jordan region, since \wp is injective on $\psi^{-n}(\alpha + Q)$. Moreover, since $f^n \circ \wp = \wp \circ \psi^n$ and since $\wp|_{\alpha + Q}$ is a homeomorphism of $\alpha + Q$ onto the 0-tile $\wp(\alpha + Q)$, the map $f^n|_{X^n}$ is a homeomorphism of the Jordan region X^n onto a 0-tile. So by Lemma 18.4 (i), the set X^n is indeed an n -tile. Since these sets X^n cover $\widehat{\mathbb{C}}$, there are no other n -tiles.

If a point $(x, y) \in \mathbb{R}^2$ (now considered as a row vector) lies in $\psi^{-n}(Q)$, then $-n2^{-n} \leq x \leq 2^{-n}$ and $0 \leq y \leq 2^{-n}$ as follows from (18.12). This implies that if we equip $\widehat{\mathbb{C}}$ with the flat metric (see Section 1.2), then for each n -tile X^n we have $\text{diam}(X^n) \lesssim n2^{-n}$. In particular,

$$\lim_{n \rightarrow \infty} \max\{\text{diam}(X^n) : X^n \text{ is an } n\text{-tile for } (f, \mathcal{C})\} = 0,$$

and so f is an expanding Thurston map. Moreover, we also conclude that $D_n := D_n(f, \mathcal{C}) \gtrsim 2^n/n$. Note that we actually have $\text{diam}(X^n) \asymp 2^{-n}/n$ for each n -tile X^n . This implies that the flat metric on $\widehat{\mathbb{C}}$ is *not* a visual metric for f (see Lemma 8.10 (ii)).

If $X_i := \wp(\psi^{-n}((0, -i) + Q))$ for $i = 1, \dots, N := \lceil 2^n/n \rceil$, then X_1, \dots, X_N is a chain of n -tiles joining opposite sides of \mathcal{C} . Hence $D_n \leq N \lesssim 2^n/n$. It follows that $D_n \asymp 2^n/n$, and so

$$\Lambda_0(f) = \lim_{n \rightarrow \infty} D_n^{1/n} = 2.$$

If Λ is the expansion factor of a visual metric, then, as we have seen in the proof Theorem 1.7 (i), we must have $1 \lesssim D_n \Lambda^{-n} \asymp 2^n \Lambda^{-n}/n$. Hence a visual metric for f with expansion factor $\Lambda = \Lambda_0(f) = 2$ does not exist.

The map f as defined above is not conformal and an example of a *Lattès-type map*. Lattès and Lattès-type maps in connection with the combinatorial expansion factor are more systematically investigated in [Yi].

19. RATIONAL THURSTON MAPS

In this section we consider rational Thurston maps, i.e., rational maps $\mathbb{R}: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ on the Riemann sphere $\widehat{\mathbb{C}}$ that are Thurston maps. Metric notions on $\widehat{\mathbb{C}}$ will usually refer to the chordal metric σ on $\widehat{\mathbb{C}}$.

We first establish a proposition that shows when a rational Thurston map is expanding. In the statement of this proposition and its proof

we will use some basic concepts of complex dynamics. For the relevant definitions and general background on this subject, see [CG] and [Mi].

Proposition 19.1. *Let $R: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ be a rational Thurston map. Then the following conditions are equivalent:*

- (i) R is expanding.
- (ii) The Julia set of R is equal to $\widehat{\mathbb{C}}$.
- (iii) R has no periodic critical points.

Proof. It suffices to establish the chain of implications (i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i).

(i) \Rightarrow (ii): Suppose R is expanding. Pick a Jordan curve $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\text{post}(R) \subset \mathcal{C}$, and let U be an arbitrary nonempty open set. Since R is expanding, the diameters of n -tiles for (f, \mathcal{C}) approach 0 uniformly as $n \rightarrow \infty$. This implies that there exist $n \in \mathbb{N}_0$, and n -tiles $X, Y \subset U$ such that X and Y are distinct, but share an n -edge. Then $R^n(X)$ and $R^n(Y)$ are the two 0-tiles. In particular, $R^n(U) = \widehat{\mathbb{C}}$. So every nonempty open set U is mapped to the whole Riemann sphere by a sufficiently high iterate of R . This implies that the Julia set \mathcal{J} of R is equal to $\widehat{\mathbb{C}}$. For otherwise, the Fatou set $\mathcal{F} = \widehat{\mathbb{C}} \setminus \mathcal{J}$ of R is a nonempty open subset of $\widehat{\mathbb{C}}$. Since \mathcal{F} is R -invariant and $\mathcal{J} \neq \emptyset$ and so $\mathcal{F} \neq \widehat{\mathbb{C}}$, we get $R^n(\mathcal{F}) = \mathcal{F} \neq \widehat{\mathbb{C}}$ for all $n \in \mathbb{N}_0$. This contradicts what we have just seen.

(ii) \Rightarrow (iii): If the Julia set of R is equal to $\widehat{\mathbb{C}}$, then its Fatou set is empty. This implies that R cannot have periodic critical points, because a periodic critical point of a rational map is part of a *super-attracting cycle* and belongs to the Fatou set.

(iii) \Rightarrow (i): Suppose R has no periodic critical points. It is a known fact from complex dynamics that then there exists a conformal metric on $\widehat{\mathbb{C}}$ with conical singularities in the points of $\text{post}(R)$ such that for the norm $\|R'(z)\|$ of the derivative with respect to this metric we have $\|R'(z)\| \geq \rho > 1$ for all $z \in \widehat{\mathbb{C}}$ (see [Mi, Thm. 19.6] or [CG, V.4.3.1]). More precisely, there exists a smooth function $\lambda: \widehat{\mathbb{C}} \setminus \text{post}(R) \rightarrow (0, \infty)$ such that for each $p \in \text{post}(R)$ we have

$$(19.1) \quad \lambda(z) \asymp \frac{1}{\sigma(z, p)^{\alpha_p}}$$

for all z near p , where $\alpha_p \in (0, 1)$. Moreover, there exists $\rho > 1$ such that

$$(19.2) \quad \|R'(z)\| = \frac{\lambda(R(z))R^\sharp(z)}{\lambda(z)} \geq \rho$$

for all $z \in \widehat{\mathbb{C}}$, where

$$R^\sharp(z) = \frac{(1 + |z|^2)|R'(z)|}{1 + |R(z)|^2}$$

is the *spherical derivative* of R and the expressions have to be understood in a suitable limiting sense at singularities (there are at most finitely many). Using λ we can introduce a path metric on $\widehat{\mathbb{C}}$ defined by

$$d(x, y) = \inf_{\gamma} \int_{\gamma} \lambda \, ds$$

for $x, y \in \widehat{\mathbb{C}}$, where integration is with respect to the spherical length element ds and the infimum is taken over all rectifiable paths $\gamma: [0, 1] \rightarrow \widehat{\mathbb{C}}$ with $\gamma(0) = x$ and $\gamma(1) = y$.

Since we have $\alpha_p \in (0, 1)$ for the exponents in (19.1), it follows that that there exist constants $C \geq 1$ and $\epsilon \in (0, 1)$ such that

$$\frac{1}{C} \sigma(z, w) \leq d(z, w) \leq C \sigma(z, w)^{1-\epsilon}$$

for all $z, w \in \widehat{\mathbb{C}}$ (actually, we can take $\epsilon = \max_{p \in \text{post}(R)} \alpha_p$). In particular, d is a metric on $\widehat{\mathbb{C}}$ that induces the standard topology on $\widehat{\mathbb{C}}$.

Moreover,

$$\text{length}_d(\gamma) = \int_{\gamma} \lambda \, ds$$

for all rectifiable paths $\gamma: [0, 1] \rightarrow \widehat{\mathbb{C}}$. This together with (19.2) implies that if $\gamma: [0, 1] \rightarrow \widehat{\mathbb{C}}$ is a path, then

$$(19.3) \quad \text{length}_d(R \circ \gamma) \geq \rho \, \text{length}_d(\gamma).$$

Now one can see that R is expanding as follows. We pick a Jordan curve $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\text{post}(R) \subset \mathcal{C}$ and consider the two 0-tiles $X_{\mathbf{w}}^0, X_{\mathbf{b}}^0 \in \mathcal{D}^0(R, \mathcal{C})$. We assume that \mathcal{C} has been chosen so that any two points in one of the 0-tiles can be joined by a path in the same tile of controlled d -length. More precisely, we require that there exists a constant L_0 such that if $x, y \in X_i^0$ for $i \in \{\mathbf{w}, \mathbf{b}\}$, then there exists a rectifiable path β in X_i^0 joining x and y such that

$$(19.4) \quad \text{length}_d(\beta) \leq L_0.$$

A Jordan curve \mathcal{C} with this property can easily be obtained if we choose it to consist of finitely many geodesic segments in the spherical metric (here we think of $\widehat{\mathbb{C}}$ as identified with the unit sphere in \mathbb{R}^3 by stereographic projection). It is clear that then we easily get a uniform estimate as in (19.4) for points x and y that stay away from the finitely

many singularities of the conformal density λ ; if one of the points is close (or even equal) to such a singularity, then one has to choose the initial piece of the connecting curve β so that its initial piece runs away from the singularity “radially”. Since $\alpha_p < 1$ for the exponent in (19.1), this leads to a favorable uniform estimate as in (19.4).

Now consider an n -tile $X \in \mathcal{D}^n(R, \mathcal{C})$, and let $u, v \in X$ be arbitrary. Then $R^n(X)$ is a 0-tile, say $R^n(X) = X_w^0$, and $R^n|X$ is a homeomorphism of X onto X_w^0 . Let $x = R^n(u)$ and $y = R^n(v)$. By choice of \mathcal{C} , we can find a path β in X_w^0 that joins x and y and satisfies $\text{length}_d(\beta) \leq L_0$. Then $\gamma := (R^n|X)^{-1} \circ \beta$ is a path in X joining u and v . We have $\beta = R^n \circ \gamma$, and so by applying (19.3) repeatedly we conclude

$$\text{length}_d(\gamma) \leq \frac{1}{\rho^n} \text{length}_d(\beta) \leq L_0 \rho^{-n}.$$

This implies $d(u, v) \leq L_0 \rho^{-n}$. Since $u, v \in X$ were arbitrary, we have

$$\text{diam}_d(X) \leq L_0 \rho^{-n}.$$

So, if as usual \mathbf{X}^n denotes the set of all n -tiles for (R, \mathcal{C}) , then

$$\max_{X \in \mathbf{X}^n} \text{diam}_d(X) \leq L_0 \rho^{-n}.$$

Since $\rho > 1$ it follows that

$$\lim_{n \rightarrow \infty} \max_{X \in \mathbf{X}^n} \text{diam}_d(X) = 0.$$

Hence R is expanding. \square

For the proof of Theorem 1.9 we need some preparation.

Let \mathcal{D} be a cell complex. A subset $\mathcal{D}' \subset \mathcal{D}$ is called a *subcomplex* of \mathcal{D} if the following condition is true: if $\tau \in \mathcal{D}'$, $\sigma \in \mathcal{D}$, and $\sigma \subset \tau$, then $\sigma \in \mathcal{D}'$. If \mathcal{D}' is a subcomplex of \mathcal{D} , then the cells in \mathcal{D}' form a cell decomposition of the underlying set

$$|\mathcal{D}'| := \bigcup \{c : c \in \mathcal{D}'\}.$$

Now suppose that $f: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ is a rational Thurston map, and $\mathcal{C} \subset \widehat{\mathbb{C}}$ is a Jordan curve with $\text{post}(f) \subset \mathcal{C}$. Consider the cell decompositions $\mathcal{D}^n = \mathcal{D}^n(f, \mathcal{C})$ of $\widehat{\mathbb{C}}$. If $\tau \in \mathcal{D}^n$, then $f^n|_\tau$ is a homeomorphism of the n -cell τ onto the 0-cell $f^n(\tau)$. So the map $\tau \mapsto f^n(\tau)$ induces a labeling $\mathcal{D}^n \rightarrow \mathcal{D}^0$ (see Definition 12.1). We call this the *natural labeling* on \mathcal{D}^n . Similarly, the map $\tau \mapsto f^n(\tau)$ induces a natural labeling on every subcomplex of \mathcal{D}^n .

Lemma 19.2. *Let $n, m \in \mathbb{N}_0$, \mathcal{D} be a subcomplex of \mathcal{D}^n , and \mathcal{D}' be a subcomplex of \mathcal{D}^m equipped with the natural labelings. If $\psi: \mathcal{D} \rightarrow \mathcal{D}'$*

is a label-preserving isomorphism, then there exists a homeomorphism $h: |\mathcal{D}| \rightarrow |\mathcal{D}'|$ such that

- (i) $h(\tau) = \psi(\tau)$ for each $\tau \in \mathcal{D}$,
- (ii) h maps $\text{int}(|\mathcal{D}|)$ conformally onto $\text{int}(|\mathcal{D}'|)$.

See Definition 12.2 for the notion of a label-preserving isomorphism of cell complexes. Roughly speaking, the lemma says that combinatorial equivalence of two subcomplexes \mathcal{D} and \mathcal{D}' gives conformal equivalence of their underlying sets.

Proof. If $\tau \in \mathcal{D}$, then $f^n(\tau) = f^m(\psi(\tau))$, because ψ is label-preserving. Hence we can define a homeomorphism $h_\tau := (f^m|_{\psi(\tau)})^{-1} \circ (f^n|_\tau)$ of τ onto $\psi(\tau)$. It is clear that the maps h_τ are compatible under inclusions of cells: if $\sigma, \tau \in \mathcal{D}$ and $\sigma \subset \tau$, then $h_\tau|_\sigma = h_\sigma$. We now define a map $h: |\mathcal{D}| \rightarrow |\mathcal{D}'|$ as follows. For $p \in |\mathcal{D}|$ pick $\tau \in \mathcal{D}$ with $p \in \tau$. Set $h(p) := h_\tau(p)$. By an argument as in the proof of Proposition 12.3 one can show that h is well-defined and a homeomorphism of $|\mathcal{D}|$ onto $|\mathcal{D}'|$. Obviously, h has property (i).

To establish property (ii) it suffices to show that h is holomorphic near each interior point p of $|\mathcal{D}|$. We consider several cases. If p is an interior point of a tile $X \in \mathcal{D}$, then this is clear by definition of h . Suppose p is an interior point of an n -edge e . Then the two n -tiles $X, Y \in \mathcal{D}^n$ that contain e in their boundaries are in \mathcal{D} . Similarly, if $X' = \psi(X)$, $Y' = \psi(Y)$, $e' = \psi(e)$, then X' and Y' are the two m -tiles that contain the m -edge e' in their boundaries. Let $X_0 = f^n(X) = f^m(X')$, $Y_0 = f^n(Y) = f^m(Y')$ and $e_0 = f^n(e) = f^m(e')$, and set $U = \text{int}(X) \cup \text{int}(e) \cup \text{int}(Y)$, $U' = \text{int}(X') \cup \text{int}(e') \cup \text{int}(Y')$, and $U_0 = \text{int}(X_0) \cup \text{int}(e_0) \cup \text{int}(Y_0)$. It follows from the considerations in the proof of Lemma 5.4 that $f^n|_U$ is a conformal map of U onto U_0 and that f^m is a conformal map of U' onto U_0 . Hence $g = (f^m|_{U'})^{-1} \circ (f^n|_U)$ is a conformal map of U onto U' . Obviously, $g = h|_U$, and so h is holomorphic near $p \in \text{int}(e) \subset U$.

Finally, if p is an n -vertex, then there is an open neighborhood $U \subset |\mathcal{D}|$ of p that contains no other n -vertex. By what we have seen, h is holomorphic on $U \setminus \{p\}$. Since h is continuous in p , the point p is a removable singularity, and hence h is holomorphic near p . \square

The following lemma provides a version of Koebe's distortion theorem for conformal maps on multiply connected regions. As we will see in the proof, the statement can easily be reduced to some classical estimates.

Lemma 19.3. *Let $\Omega \subset \widehat{\mathbb{C}}$ be a region, and $A, B \subset \Omega$ continua. Then there exist constants $C = C(A, B, \Omega) > 0$ and $C' = C'(A, \Omega) > 0$ such that for all conformal maps $h: \Omega \rightarrow \Omega' := h(\Omega)$ whose image Ω' is contained in a hemisphere of $\widehat{\mathbb{C}}$ we have*

$$(19.5) \quad \frac{1}{C} \operatorname{diam}(h(A)) \leq \operatorname{diam}(h(B)) \leq C \operatorname{diam}(h(A)),$$

and

$$(19.6) \quad \operatorname{dist}(h(A), \partial\Omega') \leq C' \operatorname{diam}(h(A)).$$

The condition that the image of h is contained in a hemisphere prevents it from being too large. It is easy to see that without such an assumption the lemma is not true in general.

Proof. Using auxiliary rotations, we may assume that $\Omega \subset \mathbb{C}$ and $\Omega' \subset \mathbb{D}$. Then the chordal and Euclidean metrics on Ω' are bi-Lipschitz equivalent. So it suffices to prove the desired inequalities for the Euclidean instead of the chordal metric.

If $h: \Omega \rightarrow \Omega'$ is conformal, and $D = D(z_0, r) \subset \Omega$ is a Euclidean disk with $2D := D(z_0, 2r) \subset \Omega$, then Koebe's distortion theorem [Po, Thm. 1.3] implies that

$$\max_{z \in D} |h'(z)| \leq C_0 \min_{z \in D} |h'(z)|,$$

where $C_0 \geq 1$ is a universal constant (one can actually take $C_0 = 81$).

If $z_1 \in A$ and $z_2 \in B$ are arbitrary, then there exists a chain D_1, \dots, D_N of Euclidean disks with $z_1 \in D_1$, $z_2 \in D_N$ and $2D_i \subset \Omega$ for $i = 1, \dots, N$, where N is bounded above by a constant only depending on A, B, Ω , but not on z_1 and z_2 . It follows that

$$\max_{z \in B} |h'(z)| \leq C_1 \min_{z \in A} |h'(z)|,$$

where C_1 only depends on A, B, Ω (and not on h). Since

$$\operatorname{diam}(h(B)) \leq C_2 \max_{z \in B} |h'(z)|,$$

and

$$(19.7) \quad \min_{z \in A} |h'(z)| \leq C_3 \operatorname{diam}(h(A)),$$

where $C_2 = C_2(B, \Omega)$ and $C_3 = C_3(A, \Omega)$, it follows that

$$\operatorname{diam}(h(B)) \leq C_4 \operatorname{diam}(h(A)),$$

where $C_4 = C_4(A, B, \Omega)$. An inequality in the opposite direction follows by reversing the roles of A and B .

By a standard estimate for conformal maps (see [Po, Cor. 1.4]) we have

$$\text{dist}(h(z), \partial\Omega') \leq |h'(z)| \text{dist}(z, \partial\Omega)$$

for all $z \in \Omega$. This implies that

$$\text{dist}(h(A), \partial\Omega') \leq C_5 \min_{z \in A} |h'(z)|,$$

where $C_5 = C_5(A)$. This together with (19.7) gives an inequality as in (19.6). \square

Proof of Theorem 1.9. Suppose first that f is an expanding Thurston map and that f is topologically conjugate to a rational map. Then this rational map is also a Thurston map, and it is expanding as one can see by an argument as in the proof of Corollary 10.6. Moreover, the conjugating map will be a snowflake equivalence with respect to visual metrics, and in particular a quasisymmetry. All this implies that we may actually assume that f itself is an expanding rational Thurston map.

By Theorem 1.2 we choose an iterate $F = f^n$ of f that has an invariant Jordan curve $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\text{post}(f) = \text{post}(F) \subset \mathcal{C}$. In the following all cells will be with respect to (F, \mathcal{C}) .

The map F is also an expanding rational Thurston map (Lemma 8.4), and the class of visual metrics for f and F are the same (Proposition 8.9 (v)). So for the desired implication of the theorem, it suffices to show that if d is any visual metric for F , then d is quasisymmetrically equivalent to the chordal metric σ .

The argument is now very similar to the considerations in [Me10], which go back to [Me02]. For the convenience of the reader we will provide full details. We will proceed in several steps. Metric notions will refer to the chordal metric.

Let $k \in \mathbb{N}_0$ and $X, Y \in \mathbf{X}^k$. Then

$$(19.8) \quad X \cap Y \neq \emptyset \Rightarrow \text{diam}(X) \asymp \text{diam}(Y).$$

Here $C(\asymp)$ is independent of X, Y , and the order k . This is seen as follows. For $k \in \mathbb{N}_0$ and non-disjoint k -tiles X and Y , consider the complex $\mathcal{D}(X, Y)$, equipped with the natural labeling, consisting of all k -cells c for which there exists a k -tile Z with $c \subset Z$ and $Z \cap (X \cup Y) \neq \emptyset$. Obviously,

$$(19.9) \quad |\mathcal{D}(X, Y)| = \bigcup \{Z \in \mathbf{X}^k : Z \cap (X \cup Y) \neq \emptyset\}.$$

Let $\Omega(X, Y)$ be the interior of $|\mathcal{D}(X, Y)|$. Then $\Omega(X, Y)$ is a region containing X and Y .

Suppose that X', Y' is a pair of non-disjoint m -tiles, $m \in \mathbb{N}_0$. We call the complexes $\mathcal{D}(X', Y')$ and $\mathcal{D}(X, Y)$ equivalent if there exists a label-preserving isomorphism $\psi: \mathcal{D}(X', Y') \rightarrow \mathcal{D}(X, Y)$ with $\psi(X') = X$ and $\psi(Y') = Y$. If $\mathcal{D}(X', Y')$ and $\mathcal{D}(X, Y)$ are equivalent, then by Lemma 19.2 there exists a conformal map $h: \Omega(X', Y') \rightarrow \Omega(X, Y)$ with $h(X') = X$ and $h(Y') = Y$.

Since F is an expanding rational Thurston map, it has no periodic critical points (Proposition 19.1) and so the length of the cycle of each vertex is uniformly bounded (see Lemma 7.2 (i) and Lemma 17.1). This implies that the number of k -tiles, and hence the number of k -cells, in $\mathcal{D}(X, Y)$ is uniformly bounded by a number independent of X, Y , and k . Therefore, among the complexes $\mathcal{D}(X, Y)$ there are only finitely many equivalence classes. Since F is expanding, there are also only finitely many complexes $\mathcal{D}(X, Y)$ such that $\Omega(X, Y)$ is not contained in a hemisphere. Hence we can find finitely many complexes $\mathcal{D}(X_1, Y_1), \dots, \mathcal{D}(X_N, Y_N)$ such that each complex $\mathcal{D}(X, Y)$ not in this list is equivalent to one complex $\mathcal{D}(X_i, Y_i)$ and such that $\Omega(X, Y)$ is contained in a hemisphere. It follows from Lemma 19.3 (applied to $A = X_i, B = Y_i, \Omega = \Omega(X_i, Y_i)$, and the conformal map $f: \Omega(X_i, Y_i) \rightarrow \Omega(X, Y)$ produced by Lemma 19.2) that $\text{diam}(X) \asymp \text{diam}(Y)$ with $C(\asymp)$ independent of X, Y , and k .

As a consequence one immediately obtains the following fact. If $m, k \in \mathbb{N}_0, X \in \mathbf{X}^k, Y \in \mathbf{X}^{k+m}$, and $Y \subset X$, then

$$(19.10) \quad \text{diam}(Y) \asymp \text{diam}(X),$$

where $C(\asymp) = C(m)$. Indeed, it is clear that $\text{diam}(Y) \leq \text{diam}(X)$. On the other hand, the number of $(m+k)$ -tiles contained in X is bounded by a number only depending on m . This and (19.8) imply that $\text{diam}(Y) \asymp \text{diam}(Z)$ whenever $Z \in \mathbf{X}^{k+m}$ and $Z \subset X$, where $C(\asymp) = C(m)$. Hence

$$\text{diam}(X) \leq \sum_{Z \in \mathbf{X}^{k+m}, Z \subset X} \text{diam}(Z) \lesssim \text{diam}(Y),$$

where $C(\lesssim) = C(m)$.

Let $k \in \mathbb{N}_0$ and $X, Y \in \mathbf{X}^k$. Then

$$(19.11) \quad X \cap Y = \emptyset \Rightarrow \text{dist}(X, Y) \gtrsim \text{diam}(X),$$

where $C(\gtrsim)$ is independent of X, Y , and k . The argument to see this is very similar to the one for (19.8).

Indeed, for $k \in \mathbb{N}_0$ and $X \in \mathbf{X}^k$ consider the cell complex $\mathcal{D}(X)$, equipped with the natural labeling, consisting of all k -cells c for which

there exists a k -tile Z with $c \subset Z$ and $Z \cap X \neq \emptyset$. Then

$$(19.12) \quad |\mathcal{D}(X)| = \bigcup \{Z \in \mathbf{X}^k : X \cap Z \neq \emptyset\}.$$

If we define $\Omega(X)$ to be the interior of $|\mathcal{D}(X)|$, then $|\mathcal{D}(X)|$ is a region containing X .

If $m \in \mathbb{N}_0$ and X' is an m -tile, then we call the complexes $\mathcal{D}(X')$ and $\mathcal{D}(X)$ equivalent if there exists a label-preserving isomorphism $\psi: \mathcal{D}(X') \rightarrow \mathcal{D}(X)$ with $\psi(X') = X$.

Again there only finitely many equivalence classes of the complexes $\mathcal{D}(X)$. Based on Lemma 19.2 and Lemma 19.3, we conclude that

$$\text{dist}(X, \partial\Omega(X)) \lesssim \text{diam}(X),$$

where $C(\lesssim)$ does not depend on X and k .

Now if Y is a k -tile with $X \cap Y = \emptyset$, then $Y \cap \Omega(X) = \emptyset$ and so

$$\text{dist}(X, Y) \leq \text{dist}(X, \partial\Omega(X)) \lesssim \text{diam}(X)$$

as desired.

Since F has no periodic critical points, the space $(\widehat{\mathbb{C}}, d)$ is doubling (Theorem 17.2). The Riemann sphere $\widehat{\mathbb{C}}$ is connected, and $(\widehat{\mathbb{C}}, \sigma)$ is also doubling. Hence, in order to establish that $\text{id}_{\widehat{\mathbb{C}}}: (\widehat{\mathbb{C}}, d) \rightarrow (\widehat{\mathbb{C}}, \sigma)$ is quasisymmetric, it is enough to show that the map $\text{id}_{\widehat{\mathbb{C}}}$ is *weakly quasisymmetric*, i.e., that there exists a constant $H \geq 1$ such that

$$(19.13) \quad d(x, y) \leq d(x, z) \Rightarrow \sigma(x, y) \leq H\sigma(x, z)$$

for all $x, y, z \in \widehat{\mathbb{C}}$ (see [He, Thm. 10.19]).

Let $m(\cdot, \cdot) = m_{F, \mathcal{C}}$ be as in Definition 8.5, and let $x, y, z \in \widehat{\mathbb{C}}$ with $d(x, y) \leq d(x, z)$ be arbitrary. We may assume that $x \neq y$. Since d is a visual metric, there exists a constant $k_0 \in \mathbb{N}$ independent of x, y, z such that

$$m(x, z) - k_0 \leq m(x, y) =: m \in \mathbb{N}_0.$$

By definition of m we can find m -tiles X and Y with $x \in X$, $y \in Y$, and $X \cap Y \neq \emptyset$. We can also find $(m + k_0 + 1)$ -tiles X' and Z' with $x \in X' \subset X$ and $z \in Z'$. Then $X' \cap Z' = \emptyset$. Thus

$$\begin{aligned} \sigma(x, y) &\leq \text{diam}(X) + \text{diam}(Y) \asymp \text{diam}(X) && \text{by (19.8)} \\ &\asymp \text{diam}(X') && \text{by (19.10)} \\ &\lesssim \text{dist}(X', Z') && \text{by (19.11)} \\ &\leq \sigma(x, z). \end{aligned}$$

Here all implicit multiplicative constants can be chosen independently of x, y, z . Hence $\text{id}_{\widehat{\mathbb{C}}}: (\widehat{\mathbb{C}}, d) \rightarrow (\widehat{\mathbb{C}}, \sigma)$ is quasisymmetric. This proves the first implication of the theorem.

For the converse direction suppose that $f: S^2 \rightarrow S^2$ is an expanding Thurston map, d is a visual metric for f on S^2 , and that there exists a quasimetry $h: (S^2, d) \rightarrow (\widehat{\mathbb{C}}, \sigma)$. Since all visual metrics are snowflake and hence also quasimetrically equivalent, we may also assume that d is a visual metric for f satisfying (1.1) in Theorem 1.7.

The map h^{-1} is also a quasimetry; so h and h^{-1} are η -quasimetric for some distortion function η . We consider the conjugate $g = h \circ f \circ h^{-1}: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ of f by h .

We claim that the family of iterates $\{g^n\}$, $n \in \mathbb{N}$, is uniformly quasiregular, i.e., each map g^n is K -quasiregular with K independent of n (see [Ri, Ch.1, Sec. 2] for the definition of a K -quasiregular map). The reason is that with the metric d satisfying (1.1), the map f is locally “conformal”, and so the dilatation of $g^n = h \circ f^n \circ h^{-1}$ can be bounded by the dilatations of h and h^{-1} , and hence by a constant independent of n .

To be more precise, let $n \in \mathbb{N}$, $u \in \widehat{\mathbb{C}}$, and for small $\epsilon > 0$ consider points $v, w \in \widehat{\mathbb{C}}$ with $\sigma(u, v) = \sigma(u, w) = \epsilon$. Define $x = h^{-1}(u)$, $y = h^{-1}(v)$, $z = h^{-1}(w)$. By Theorem 1.7 we have that if $\epsilon > 0$ is sufficiently small (depending on u and n), then

$$\begin{aligned} \frac{\sigma(g^n(u), g^n(v))}{\sigma(g^n(u), g^n(w))} &= \frac{\sigma(h(f^n(x)), h(f^n(y)))}{\sigma(h(f^n(x)), h(f^n(z)))} \\ &\leq \eta \left(\frac{d(f^n(x), f^n(y))}{d(f^n(x), f^n(z))} \right) = \eta \left(\frac{d(x, y)}{d(x, z)} \right) \\ &= \eta \left(\frac{d(h^{-1}(u), h^{-1}(v))}{d(h^{-1}(u), h^{-1}(w))} \right) \leq H := \eta(\eta(1)). \end{aligned}$$

Hence

$$H(g^n, u) := \limsup_{\epsilon \rightarrow 0} \max \left\{ \frac{\sigma(g^n(u), g^n(v))}{\sigma(g^n(u), g^n(w))} : v, w \in \widehat{\mathbb{C}}, \sigma(u, v) = \sigma(u, w) = \epsilon \right\} \leq H$$

for all $u \in \widehat{\mathbb{C}}$ and $n \in \mathbb{N}$. This inequality implies that g^n is locally H -quasiconformal on the set $\widehat{\mathbb{C}} \setminus \text{crit}(g^n)$ (according to the so-called “metric” definition of quasiconformality; see [Vä, Sect. 34]).

In particular, this implies that $g^n|_{\widehat{\mathbb{C}} \setminus \text{crit}(g^n)}$ is K -quasiregular with $K = K(H)$ independent of n . Since the finite set $\text{crit}(g^n)$ is removable for quasiregularity (see [Ri, Ch. 7, Sect. 1]), we conclude that g^n is K -quasiregular with K independent of n .

So the family of iterates $\{g^n\}$ of g is uniformly quasiregular. This implies that g is topologically conjugate to a rational map (see [IM,

p. 508, Thm. 21.5.2]). Hence f is also topologically conjugate to a rational map. \square

In the previous proof we actually showed the following fact.

Corollary 19.4. *Let $R: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ be an expanding rational Thurston map. Then every visual metric for R on $\widehat{\mathbb{C}}$ is quasimetrically equivalent to the chordal metric on $\widehat{\mathbb{C}}$.*

Our previous results now immediately give a proof of Theorem 1.1.

Proof of Theorem 1.1. Let $f: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ be a Thurston map whose Julia set is the whole Riemann sphere. Then f is expanding (Proposition 19.1) and so by Theorem 1.2 there exist a Jordan curve $\mathcal{C} \subset \widehat{\mathbb{C}}$ with $\text{post}(f) \subset \mathcal{C}$ that is invariant for some iterate f^n of f . By Theorem 1.8 the curve \mathcal{C} is a quasicircle if it is equipped with a visual metric for f . Corollary 19.4 implies that \mathcal{C} is also a quasicircle in the usual sense, i.e., a quasicircle if equipped with the chordal metric on $\widehat{\mathbb{C}}$. \square

Proof of Theorem 1.10. Suppose that $f: S^2 \rightarrow S^2$ is a Thurston map with $\#\text{post}(f) = 3$. We pick a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$ and consider the cell decompositions $\mathcal{D}^0 := \mathcal{D}^0(f, \mathcal{C})$ and $\mathcal{D}^1 := \mathcal{D}^1(f, \mathcal{C})$. Since $\#\text{post}(f) = 3$, the tiles in \mathcal{D}^0 and \mathcal{D}^1 are *(topological) triangles*, i.e., they contain three vertices and edges in their boundaries.

Let Δ be a fixed equilateral Euclidean triangle of side-length 1. There exists a (non-unique) path metric d_1 on S^2 such that each tile X in \mathcal{D}^1 equipped with d_1 is isometric to Δ (here and below it is understood that the vertices of the triangles correspond to each other under the isometry). Then, roughly speaking, the metric space (S^2, d_1) is an (abstract) polyhedral surface that can be obtained by gluing together different copies of Δ , one for each tile in \mathcal{D}^1 , according to the combinatorics of \mathcal{D}^1 .

Similarly, there is a path metric d_0 on S^2 such that each of the two tiles in \mathcal{D}^0 equipped with d_0 is isometric to Δ . The metric space (S^2, d_0) is isometric to a “pillow” consisting of two copies of Δ glued together along their boundaries.

The map f induces an orientation-preserving labeling $c \in \mathcal{D}^1 \mapsto f(c) \in \mathcal{D}^0$. Note that if X is a tile in \mathcal{D}^1 , then X and $f(X)$ if equipped with d_1 and d_0 , respectively, are both isometric to Δ , and hence to each other. We can arrange this isometry so that it agrees with the map f on the vertices of X . By using these piecewise isometries on tiles, one can construct a unique map $g: S^2 \rightarrow S^2$ that is cellular for $(\mathcal{D}^1, \mathcal{D}^0)$ and compatible with the given labeling, and has the property that g restricted to any cell $c \in \mathcal{D}^1$ equipped with the metric d_1 is an isometry

onto the corresponding cell $g(c) = f(c) \in \mathcal{D}^0$ equipped with d_0 . Then $\text{post}(f) = \text{post}(g)$, and the maps f and g are Thurston equivalent by Proposition 12.5.

The surface (S^2, d_1) carries a unique conformal structure compatible with the polyhedral structure (away from the vertices, one uses orientation-preserving local isometries to the Euclidean plane as chart maps on the surface; this determines the conformal structure uniquely, because each of the finitely many vertices is a “removable singularity”). By the uniformization theorem there exists a conformal map α from (S^2, d_1) (considered as a Riemann surface) onto the Riemann sphere $\widehat{\mathbb{C}}$. Similarly, (S^2, d_0) has a natural Riemann surface structure and there is a conformal map $\beta: (S^2, d_0) \rightarrow \widehat{\mathbb{C}}$. By post-composing β by suitable Möbius transformation if necessary, we may assume that the maps α and β are identical on the 3-element set $P = \text{post}(f) = \text{post}(g)$. Since α and β are orientation-preserving, these homeomorphisms are then isotopic rel. P by Lemma 10.11.

The map $g: (S^2, d_1) \rightarrow (S^2, d_0)$ is a local isometry away from the vertices of \mathcal{D}^1 . So this map is a holomorphic map from the Riemann surface (S^2, d_1) to the Riemann surface (S^2, d_0) . Hence $R := \beta \circ g \circ \alpha^{-1}$ is a holomorphic map from $\widehat{\mathbb{C}}$ to $\widehat{\mathbb{C}}$, and so a rational map. Moreover, R is a Thurston map (this follows from the argument used to establish (3.4)) and R is Thurston equivalent to g and hence to f . The first part of the theorem follows.

Suppose in addition that f is expanding and has no periodic critical points. Since the latter condition is invariant under Thurston equivalence, the rational Thurston map R constructed as above will then not have periodic critical points either, and is hence expanding by Proposition 19.1. Therefore, the maps f and R are topologically conjugate by Theorem 10.4.

Conversely, if f is expanding and topologically conjugate to a rational map R , then R is an expanding Thurston map. Hence R has no periodic critical points by Proposition 19.1, which implies that f cannot have periodic critical points either. \square

20. THE MEASURE OF MAXIMAL ENTROPY

In this section we investigate measure-theoretic properties of the dynamical system given by iteration of a Thurston map. We first review some facts about topological and metric entropy. For more background on these topics see [KH, Wa].

In the following, (X, d) is a compact metric space, and $g: X \rightarrow X$ a continuous map. For $n \in \mathbb{N}_0$, and $x, y \in X$ we define

$$(20.1) \quad d_g^n(x, y) = \max\{d(g^k(x), g^k(y)) : k = 0, \dots, n-1\}.$$

Then d_g^n is a metric on X . Let $D(g, \epsilon, n)$ be the minimal number of sets whose d_g^n -diameter is at most $\epsilon > 0$ and whose union covers X . One can show that the limit

$$h(g, \epsilon) := \lim_{n \rightarrow \infty} \frac{1}{n} \log(D(g, \epsilon, n))$$

exists [KH, p. 109, Lem. 3.1.5]. Obviously, the quantity $h(g, \epsilon)$ is non-increasing in ϵ . One defines the *topological entropy* of g (see [KH, Sect. 3.1.b]) as

$$h_{\text{top}}(g) := \lim_{\epsilon \rightarrow 0} h(g, \epsilon) \in [0, \infty].$$

If one uses another metric d' on X , then one obtains the same quantity for $h_{\text{top}}(g)$ if d' induces the same topology on X as d [KH, p. 109, Prop. 3.1.2]. The topological entropy is also well-behaved under iteration. Indeed, if $n \in \mathbb{N}$, then $h_{\text{top}}(g^n) = nh_{\text{top}}(g)$ [KH, p. 111, Prop. 3.1.7 (3)].

We denote by \mathcal{B} the σ -algebra of all Borel sets on X . A measure on X is understood to be a Borel measure, i.e., one defined on \mathcal{B} . If X is compact and a measure μ is called *g -invariant* if

$$(20.2) \quad \mu(g^{-1}(A)) = \mu(A)$$

for all $A \in \mathcal{B}$. Note that by continuity of g , we have $g^{-1}(A) \in \mathcal{B}$ whenever $A \in \mathcal{B}$. We denote by $\mathcal{M}(X, g)$ the set of all g -invariant Borel probability measures on X .

If μ is a probability measure on a compact metric space X , then it is *regular*. This means that for every $\epsilon > 0$ and every Borel set $A \subset X$ there exists a compact set $K \subset A$ with $\mu(A \setminus K) < \epsilon$ (*inner regularity*) and an open set $U \subset X$ with $A \subset U$ and $\mu(U \setminus A) < \epsilon$ (*outer regularity*). See [Ru, Thm. 2.18] for a more general result that contains this statement as a special case.

A *semi-algebra* \mathcal{S} is a system of sets in X satisfying the following properties: (i) $\emptyset \in \mathcal{S}$, (ii) $A \cap B \in \mathcal{S}$, whenever $A, B \in \mathcal{S}$, and (iii) $X \setminus A$ is a finite union of disjoint sets in \mathcal{S} , whenever $A \in \mathcal{S}$. A semi-algebra *generates* \mathcal{B} if \mathcal{B} is the smallest σ -algebra containing \mathcal{S} .

Let \mathcal{S} be a semi-algebra generating \mathcal{B} . If μ and ν are two measures on X and $\mu(A) = \nu(A)$ for all $A \in \mathcal{S}$, then $\mu = \nu$. Similarly, in order to show that a measure μ is g -invariant it is enough to verify (20.2) for all sets A in \mathcal{S} (see [Wa, p. 20, proof of Thm. 1.1] for the simple argument on how to prove these statements).

Let $\mu \in \mathcal{M}(X, g)$. Then g is called *ergodic* for μ if for each set $A \in \mathcal{B}$ with $g^{-1}(A) = A$ we have $\mu(A) = 0$ or $\mu(A) = 1$. The map g is called *mixing* for μ if

$$(20.3) \quad \lim_{n \rightarrow \infty} \mu(g^{-n}(A) \cap B) = \mu(A)\mu(B)$$

for all $A, B \in \mathcal{B}$. Obviously, if g is mixing for μ , then g is also ergodic. To establish mixing, one only has to verify (20.3) for sets A and B in a semi-algebra generating \mathcal{B} ([Wa, p. 41, Thm. 1.17 (iii)]; note that the terminology in [Wa] slightly differs from ours). If $\mu, \nu \in \mathcal{M}(X, g)$, g is ergodic for μ , and ν is absolutely continuous with respect to μ , then $\nu = \mu$ [Wa, p. 153, Rems. (1)].

Our next goal is to define the metric entropy of g for a measure μ . We will follow [KH, Sect. 4.3] with slight differences in notation and terminology (see also [Wa, Ch. 4]).

Let $\mu \in \mathcal{M}(X, g)$. A *measurable partition* ξ for (X, μ) is a countable collection $\xi = \{A_i : i \in I\}$ of sets in \mathcal{B} such that $\mu(A_i \cap A_j) = 0$ for $i, j \in I$, $i \neq j$ and

$$\mu\left(X \setminus \bigcup_{i \in I} A_i\right) = 0.$$

Here I is a countable (i.e., finite or countably infinite) index set. The *symmetric difference* of two sets $A, B \subset X$ is defined as

$$A \Delta B = (A \setminus B) \cup (B \setminus A).$$

Two measurable partitions ξ and η for (X, μ) are called *equivalent* if there exists a bijection between the sets of positive measure in ξ and the sets of positive measure in η such that corresponding sets have a symmetric difference of vanishing μ measure. Roughly speaking, this means that the partitions are the same up to sets of measure zero.

Let $\xi = \{A_i : i \in I\}$ and $\eta = \{B_j : j \in J\}$ be measurable partitions of (X, μ) . Then

$$\xi \vee \eta := \{A_i \cap B_j : i \in I, j \in J\}$$

is also a measurable partition, called the *join* of ξ and η . The join of finitely many measurable partitions is defined similarly.

Let

$$g^{-1}(\xi) := \{g^{-1}(A_i) : i \in I\}$$

and for $n \in \mathbb{N}$ define

$$\xi_g^n = \xi \vee g^{-1}(\xi) \vee \dots \vee g^{-(n-1)}(\xi).$$

The *entropy* of ξ (for given g) is

$$H_\mu(g, \xi) = \sum_{i \in I} \mu(A_i) \log(1/\mu(A_i)) \in [0, \infty].$$

Here it is understood that the function $\phi(x) = x \log(1/x)$ is continuously extended to 0 by setting $\phi(0) = 0$.

One can show that if $H_\mu(g, \xi) < \infty$, then the quantities $H_\mu(g, \xi_g^n)$, $n \in \mathbb{N}_0$, are *subadditive* in the sense that

$$H_\mu(g, \xi_g^{n+k}) \leq H_\mu(g, \xi_g^n) + H_\mu(g, \xi_g^k)$$

for all $k, n \in \mathbb{N}_0$ [KH, p. 168, Prop. 4.3.6]. This implies ([Wa, p. 87, Thm. 4.9]; see also the argument in the last part of the proof of Lemma 11.3) that

$$h_\mu(g, \xi) := \lim_{n \rightarrow \infty} \frac{1}{n} H_\mu(g, \xi_g^n) \in [0, \infty)$$

exists and we have

$$h_\mu(g, \xi) = \inf_{n \in \mathbb{N}} \frac{1}{n} H_\mu(g, \xi_g^n).$$

The quantity $h_\mu(g, \xi)$ is called the *(metric) entropy of g relative to ξ* . The *(metric) entropy of g for μ* is defined as

$$(20.4) \quad h_\mu(g) = \sup \{ h_\mu(g, \xi) : \xi \text{ is a measurable partition of } (X, \mu) \text{ with } H_\mu(g, \xi) < \infty \}.$$

In this definition it is actually enough to take the supremum over all finite measurable partitions ξ (this easily follows from “Rokhlin’s inequality” [KH, p. 169, Prop. 4.3.10 (4)]).

We call a finite measurable partition ξ a *generator* for (g, μ) if the following condition is true: Let \mathcal{A} be the smallest σ -algebra containing all sets in the partitions ξ_g^n , $n \in \mathbb{N}$. Then for each Borel set $B \in \mathcal{B}$ there exists a set $A \in \mathcal{A}$ such that $\mu(A \Delta B) = 0$. If for every set $B \in \mathcal{B}$ and for every $\epsilon > 0$, there exists $n \in \mathbb{N}$ and a union A of sets in ξ_g^n with $\mu(A \Delta B) < \epsilon$, then ξ is a generator for (g, μ) . If ξ is a generator, then $h_\mu(g) = h_\mu(g, \xi)$ by the Kolmogorov-Sinai Theorem [Wa, p. 95, Thm. 4.17].

If $\mu \in \mathcal{M}(g, X)$ and $n \in \mathbb{N}$, then [KH, pp. 171–172, Prop. 4.3.16 (4)]

$$h_\mu(g^n) = n h_\mu(g).$$

If $\alpha \in [0, 1]$ and $\nu \in \mathcal{M}(g, X)$ is another measure, then [Wa, p. 183, Thm. 8.1]

$$h_{\alpha\mu + (1-\alpha)\nu}(g) = \alpha h_\mu(g) + (1 - \alpha) h_\nu(g).$$

The topological entropy is related to the metric entropy by the so-called variational principle. It states that [Wa, p. 188, Thm. 8.6]

$$h_{top}(g) = \sup\{h_\mu(g) : \mu \in \mathcal{M}(g, X)\}.$$

A measure $\mu \in \mathcal{M}(g, X)$ for which $h_{top}(g) = h_\mu(g)$ is called a *measure of maximal entropy*.

Let \tilde{X} be another compact metric space. If μ is a measure on X and $\varphi: X \rightarrow \tilde{X}$ is continuous, then the *push-forward* $\varphi_*\mu$ of μ by φ is the measure given by $\varphi_*\mu(A) = \mu(\varphi^{-1}(A))$ for all Borel sets $A \subset \tilde{X}$.

Suppose $\tilde{g}: \tilde{X} \rightarrow \tilde{X}$ is a continuous map, and $\mu \in \mathcal{M}(X, g)$ and $\tilde{\mu} \in \mathcal{M}(\tilde{X}, \tilde{g})$. Then the dynamical system $(\tilde{X}, \tilde{g}, \tilde{\mu})$ is called a (*topological*) *factor* of (X, g, μ) if there exists a continuous map $\varphi: X \rightarrow \tilde{X}$ such that $\varphi_*\mu = \tilde{\mu}$ and $\tilde{g} \circ \varphi = \varphi \circ g$. In this case $h_{\tilde{\mu}}(\tilde{g}) \leq h_\mu(g)$ [KH, p. 171, Prop. 4.3.16].

Now let S^2 be a 2-sphere and $f: S^2 \rightarrow S^2$ be an expanding Thurston map. Our goal is to describe a measure of maximal entropy for f and show its uniqueness. Existence and uniqueness of such a measure can also be derived from work by Pilgrim and Haïssinsky [HP09] who used the so-called thermodynamical formalism for this purpose. We will present a direct elementary argument that has the advantage of giving additional insight into the dynamical behavior of f .

By Theorem 1.2 we can fix a sufficiently high iterate $F = f^n$ of f that has an F -invariant Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) = \text{post}(F) \subset \mathcal{C}$. Then F is also an expanding Thurston map (Lemma 8.4). In the following we consider the cell decompositions $\mathcal{D}^k = \mathcal{D}^k(F, \mathcal{C})$ for $k \in \mathbb{N}_0$. As usual a *cell* is a cell in any of the cell decompositions \mathcal{D}^k , $k \in \mathbb{N}_0$, and the terms *tiles*, *edges*, and *vertices* are used in a similar way.

By Proposition 11.1 the cell decomposition \mathcal{D}^{m+k} is a refinement of \mathcal{D}^k for $m, k \in \mathbb{N}_0$, and cells are subdivided by cells of higher order.

We denote by $X_{\mathbf{w}}^0$ and $X_{\mathbf{b}}^0$ the two 0-tiles, and color the tiles for (F, \mathcal{C}) as in Lemma 6.2. In particular, $X_{\mathbf{w}}^0$ is colored white and $X_{\mathbf{b}}^0$ is colored black.

For $k \in \mathbb{N}_0$ let w_k be the number of white and b_k be the number of black k -tiles contained in $X_{\mathbf{w}}^0$, and similarly let w'_k and b'_k be the number of white and black k -tiles contained in $X_{\mathbf{b}}^0$. Then considerations as in the last part of the proof of Lemma 5.4 imply that

$$(20.5) \quad w_k + w'_k = b_k + b'_k = \deg(F)^k.$$

Note that $b_1, w'_1 \neq 0$. Indeed, suppose that $b_1 = 0$, for example. Then $X_{\mathbf{w}}^0$ contains only white 1-tiles. Let $X \subset X_{\mathbf{w}}^0$ be such a 1-tile, $e \subset X$ be a 1-edge with $e \subset \partial X$, and Y be the other 1-tile containing

e. Then Y is black and so $Y \subset X_{\mathbf{b}}^0$. Hence

$$e \subset X \cap Y \subset X_{\mathbf{w}}^0 \cap X_{\mathbf{b}}^0 = \partial X_{\mathbf{w}}^0.$$

Since ∂X is a union of 1-edges, it follows that $\partial X \subset \partial X_{\mathbf{w}}^0$. As $X_{\mathbf{w}}^0$ and X are Jordan regions and $X \subset X_{\mathbf{w}}^0$, this is only possible if $X = X_{\mathbf{w}}^0$. Hence $X_{\mathbf{w}}^0$ is a 1-tile and $F|X_{\mathbf{w}}^0$ is a homeomorphism of $X_{\mathbf{w}}^0$ onto itself. Applying Lemma 18.4 (i) repeatedly, it follows that $X_{\mathbf{w}}^0$ is a k -tile for each $k \in \mathbb{N}_0$. This is impossible, because F is expanding and so the diameters of k -tiles approach 0 as $k \rightarrow \infty$.

Define

$$(20.6) \quad w := \frac{b_1}{b_1 + w'_1}, \quad b := \frac{w'_1}{b_1 + w'_1}.$$

Then $w, b > 0$ and $w + b = 1$. It follows from (20.5) for $k = 1$ that the matrix

$$(20.7) \quad A = \begin{pmatrix} w_1 & b_1 \\ w'_1 & b'_1 \end{pmatrix}$$

has the eigenvalues $\lambda_1 = \deg(F)$ and $\lambda_2 = w_1 - b_1$ with respective eigenvectors

$$v_1 = \begin{pmatrix} w \\ b \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Here $|\lambda_2| = |w_1 - b_1| < \lambda_1 = \deg(F)$. Indeed, since $1 \leq b_1 \leq \deg(F)$ and $0 \leq w_1 \leq \deg(F)$, we otherwise had $w_1 = 0$ and $b_1 = \deg(F) \geq 2$. Then the white 0-tile contains only black 1-tiles. Arguing similarly as in the discussion before the lemma, there can then be only one such tile, and so $b_1 = 1$. This is a contradiction.

The existence of a largest positive eigenvalue λ_1 for A with a corresponding eigenvector with all positive coordinates is an instance of the Perron-Frobenius Theorem ([KH, p. 52, Thm. 1.9.11]).

Let $k, l, m \in \mathbb{N}_0$ with $m \geq l \geq k$ be arbitrary. The map F^k preserves colors of tiles, i.e., if X^m is an m -tile, then $F^k(X^m)$ is an $(m - k)$ -tile with the same color as X^m . Moreover, if Y^l is an l -tile, then it follows from Lemma 18.4 (i) that the map $X^m \mapsto F^k(X^m)$ induces a bijection between the m -tiles contained in Y^l and the $(m - k)$ -tiles contained in the $(l - k)$ -tile $Y^{l-k} := F^k(Y^l)$.

If we use this for $m = k + 1$ and $l = k$, then we see that a white k -tile contains w_1 white and b_1 black $(k + 1)$ -tiles, and similarly each black k -tile contains w'_1 white and b'_1 black $(k + 1)$ -tiles. This leads to the identity

$$(20.8) \quad \begin{pmatrix} w_{k+1} & b_{k+1} \\ w'_{k+1} & b'_{k+1} \end{pmatrix} = \begin{pmatrix} w_k & b_k \\ w'_k & b'_k \end{pmatrix} \begin{pmatrix} w_1 & b_1 \\ w'_1 & b'_1 \end{pmatrix}.$$

for $k \in \mathbb{N}_0$. The following lemma is a consequence.

Lemma 20.1. *For all $k \in \mathbb{N}$ we have*

$$\begin{aligned} w_k &= w \deg(F)^k + b(w_1 - b_1)^k, & b_k &= w \deg(F)^k - w(w_1 - b_1)^k, \\ w'_k &= b \deg(F)^k - b(w_1 - b_1)^k, & b'_k &= b \deg(F)^k + w(w_1 - b_1)^k. \end{aligned}$$

Since $|w_1 - b_1| < \deg(F)$, the terms with $\deg(F)^k$ in these identities are the main terms for large k .

Proof. This follows from (20.8) by induction. \square

The following lemma will be instrumental in proving that each edge is a set of measure zero for the measure of maximal entropy of F .

Lemma 20.2. *There exists $1 \leq L < \deg(F)$ such that for all $k, m \in \mathbb{N}_0$ and each m -edge e there exists a collection M of $(m+k)$ -tiles with $\#M \leq CL^k$ such that e is contained in the interior of the set $\bigcup_{X \in M} X$. Here C is independent of k .*

The total number of $(m+k)$ -tiles is $2 \deg(F)^{m+k}$. So the lemma says that for large k , the m -edge e can be covered by a substantially smaller number of $(m+k)$ -tiles.

Proof. It follows from Lemma 8.13 and Lemma 8.10 (i) that one can find $k_0 \in \mathbb{N}$ such that for every s -tile X , $s \in \mathbb{N}_0$, there exist two $(s+k_0)$ -tiles Y and Z , one white and one black, with $Y \subset \text{int}(X)$ and $Z \subset \text{int}(X)$.

Every white s -tile contains w_{k_0} white and b_{k_0} black $(s+k_0)$ -tiles, and every black s -tile contains w'_{k_0} white and b'_{k_0} black $(s+k_0)$ -tiles. By (20.5) we also know that $w_{k_0} + w'_{k_0} = b_{k_0} + b'_{k_0} = \deg(F)^{k_0}$.

Now let e be an arbitrary m -edge. For each $l \in \mathbb{N}_0$ we will define certain collections of $(m+lk_0)$ -tiles T_l whose union contains e . We will denote the number of white tiles in T_l by N_l^w , the number of black tiles in T_l by N_l^b , and define $N_l = \max\{N_l^w, N_l^b\}$. Then the number of tiles in T_l is bounded by $2N_l$.

Let T_0 be the set of all m -tiles that meet e . Then the union of the tiles in T_0 is the closure of the edge flower of e and so it contains e in its interior.

Suppose the collection T_l has been constructed. Then we subdivide each of the tiles U in T_l into $(m+(l+1)k_0)$ -tiles and remove one white and one black tile $(m+(l+1)k_0)$ -tile from the interior of U . We define T_{l+1} as the collection of all tiles obtained in this way from tiles in T_l . Since $\text{int}(U) \cap e = \emptyset$ for each $U \in T_l$, the union of the tiles in T_{l+1} still

contains e in its interior. Then for the number of white tiles in T_{l+1} we have the estimate

$$\begin{aligned} N_{l+1}^{\mathbf{w}} &= N_l^{\mathbf{w}}(w_{k_0} - 1) + N_l^{\mathbf{b}}(w'_{k_0} - 1) \\ &\leq N_l(w_{k_0} + w'_{k_0} - 2) = N_l(\deg(F)^{k_0} - 2). \end{aligned}$$

Similarly,

$$N_{l+1}^{\mathbf{b}} \leq N_l(\deg(F)^{k_0} - 2),$$

and so

$$N_{l+1} \leq N_l(\deg(F)^{k_0} - 2).$$

Let

$$L := (\deg(F)^{k_0} - 2)^{1/k_0} < \deg(F).$$

Then

$$\#T_l \leq 2N_l \leq 2N_0 L^{k_0 l}$$

is a bound for the total number of tiles in T_l .

Now let $k \in \mathbb{N}$ be arbitrary. Let $l \in \mathbb{N}_0$ be the smallest number with $lk_0 \geq k$. For each tile $(m + lk_0)$ -tile U in T_l we can pick a $(m + k)$ -tile that contains U . Let M be that collection of all $(m + k)$ -tiles obtained in this way. Then the union of all tiles in M contains e in its interior and we have

$$\#M \leq \#T_l \leq 2N_0 L^{k_0 l} \leq 2N_0 L^{k_0} L^k = CL^k,$$

where $C = 2N_0 L^{k_0}$. The claim follows. \square

The constant C in the previous lemma depends on e . If we require the weaker property that the collection M of $(m + k)$ -tiles only covers e , then we can choose the collection so that $\#M \leq CL^k$ with a constant C independent of e . Indeed, in this case, we can choose T_0 to consist of the two m -tiles X and Y , one white and one black, that contain e in their boundary. Then $N_0 = 1$ and this leads to an inequality of the desired type with a constant C independent of e .

In the next lemma we obtain an upper bound for the topological entropy of f .

Lemma 20.3. $h_{\text{top}}(f) \leq \log(\deg(f)).$

As we will see later, we actually have $h_{\text{top}}(f) = \log(\deg(f))$ (see Corollary 20.8).

Proof. Since $h_{\text{top}}(F) = nh_{\text{top}}(f)$ and $\deg(F) = \deg(f)^n$, it suffices to show that $h_{\text{top}}(F) \leq \log(\deg(F))$.

To show that $h_{\text{top}}(F) \leq \log(\deg(F))$, we fix a base metric d on S^2 and let $\epsilon > 0$ be arbitrary. Since F is expanding, we can find $k_0 \in \mathbb{N}_0$ such that $\text{diam}(X) \leq \epsilon$ whenever $X \in \mathbf{X}^k$ for $k \geq k_0$.

Now if $k \in \mathbb{N}_0$ and $X \in \mathbf{X}^{k+k_0}$ is arbitrary, then $F^i(X)$ is a $(k-i+k_0)$ -tile for $i = 0, 1, \dots, k$, and so $\text{diam}(F^i(X)) < \epsilon$. This implies that the diameter of X with respect to the metric d_F^k (see (20.1)) is $\leq \epsilon$. Since the number of $(k+k_0)$ -tiles is equal to $2 \deg(F)^{k+k_0}$ and these tiles form a cover of S^2 , it follows that $D(F, \epsilon, k) \leq 2 \deg(F)^{k+k_0}$, and so $h(\epsilon, F) \leq \log(\deg(F))$. Letting $\epsilon \rightarrow 0$ we conclude $h_{\text{top}}(F) \leq \log(\deg(F))$ as desired. \square

Since the curve \mathcal{C} is F -invariant, we can restrict F to \mathcal{C} to obtain a map $F|_{\mathcal{C}}: \mathcal{C} \rightarrow \mathcal{C}$. The following lemma shows that the topological entropy of this restriction is strictly smaller than $\log(\deg(F))$.

Lemma 20.4. $h_{\text{top}}(F|_{\mathcal{C}}) < \log(\deg(F))$.

Proof. The proof is very similar to the first part of the proof of Lemma 20.3. Again let d be a base metric on S^2 .

Since \mathcal{C} consists of $\# \text{post}(f)$ 0-edges, by Lemma 20.2 we can cover \mathcal{C} by a collection M_k of k -tiles, where $\#M_k \leq CL^k$. Here $1 \leq L < \deg(F)$ and C is independent of k . The k -edges in the boundaries of the k -tiles in M_k then form a cover of \mathcal{C} . It is clear that each k -edge contained in \mathcal{C} must belong to this collection. Hence if E_k is the set of all k -edges contained in \mathcal{C} , we have $\#E_k \leq C'L^k$ with a constant C' independent of k .

Now let $\epsilon > 0$ be arbitrary. Since F is expanding, we can find $k_0 \in \mathbb{N}_0$ such that $\text{diam}(X) \leq \epsilon$ whenever $X \in \mathbf{X}^k$ for $k \geq k_0$. Since every k -edge is contained in a k -tile, we also have $\text{diam}(e) \leq \epsilon$ whenever $e \in \mathbf{E}^k$ for $k \geq k_0$.

If $k \in \mathbb{N}_0$ and $e \in E_{k+k_0}$ is arbitrary, then $F^i(e)$ is a $(k-i+k_0)$ -edge for $i = 0, 1, \dots, k$, and so $\text{diam}(F^i(e)) < \epsilon$. This implies that the diameter of e with respect to the metric d_F^k is $\leq \epsilon$.

It follows that $D(F|_{\mathcal{C}}, \epsilon, k) \leq \#E_{k+k_0} \leq C'L^{k_0+k}$, and so $h(\epsilon, F) \leq \log(L)$. Letting $\epsilon \rightarrow 0$ we conclude $h_{\text{top}}(F) \leq \log(L) < \log(\deg(F))$ as desired. \square

In the following we let

$$E^\infty = \bigcup_{k \in \mathbb{N}_0} F^{-k}(\mathcal{C}).$$

Then E^∞ is a Borel set. Proposition 6.1 (iii) (applied to the map F) implies that E^∞ is equal to union of all edges. Since every vertex is contained in an edge, the set E^∞ also contains all vertices. Moreover, we have

$$(20.9) \quad F^{-1}(E^\infty) = E^\infty.$$

Indeed, since the preimage of every edge is a union of edges, it is clear that $F^{-1}(E^\infty) \subset E^\infty$. For the other inclusion note that $F(\mathcal{C}) \subset \mathcal{C}$ and so

$$F(E^\infty) \subset \bigcup_{k \in \mathbb{N}_0} F(F^{-k}(\mathcal{C})) \subset E^\infty \cup F(\mathcal{C}) = E^\infty.$$

Lemma 20.5. *Let μ be an F -invariant probability measure on S^2 with $\mu(E^\infty) = 0$. Then for each $k \in \mathbb{N}$ the set \mathbf{X}^k of k -tiles forms a measurable partition of S^2 . It is equivalent to the partition ξ_F^k where $\xi = \mathbf{X}^1$. Moreover, $\xi = \mathbf{X}^1$ is a generator for (F, μ) .*

Proof. Note that $\mu(E^\infty) = 0$ implies that all edges are sets of μ -measure zero. Since every vertex is contained in an edge, we also have $\mu(\{v\}) = 0$ for all vertices v . The k -tiles cover S^2 and two distinct k -tiles have only edges or vertices, i.e., a set of μ -measure zero in common. Hence \mathbf{X}^k is a measurable partition of S^2 .

Let $X \in \mathbf{X}^k$ be arbitrary. Then for $i = 1, \dots, k$ there exist unique i -tiles X_i with $X = X_k \subset X_{k-1} \subset \dots \subset X_1$. Put $Y_i = F^{i-1}(X_i)$ for $i = 1, \dots, k$. Then Y_1, \dots, Y_k are 1-tiles. We claim that

$$(20.10) \quad X = Y_1 \cap F^{-1}(Y_2) \cap \dots \cap F^{-(k-1)}(Y_k).$$

To verify this, denote the right hand side in this equation by \tilde{X} . Then it is clear that $X \subset \tilde{X}$. We verify $X = \tilde{X}$ by inductively showing that for any point $x \in \tilde{X}$ we have $x \in X_i$ for $i = 1, \dots, k$, and so $x \in X_k = X$.

Indeed, since $\tilde{X} \subset Y_1 = X_1$ this is clear for $i = 1$. Suppose $x \in X_i$ for some $i < k$. Then $F^i|_{X_i}$ is a homeomorphism of X_i onto the 0-tile $F^i(X_i)$. Moreover, $x \in X_i$, $X_{i+1} \subset X_i$ and $F^i(x) \in Y_{i+1} = F^i(X_{i+1})$. Hence by injectivity of F^i on X_i we have $x \in X_{i+1}$.

Equation (20.10) shows that every element in \mathbf{X}^k belongs to ξ_F^k . This implies that the measurable partitions \mathbf{X}^k and ξ_F^k are equivalent (ξ_F^k may contain additional sets, but they have to be of measure zero).

To establish that ξ is a generator, let $B \subset S^2$ be an arbitrary Borel set and $\epsilon > 0$. By what we have seen, it is enough to show that there exists $k \in \mathbb{N}$ and a union A of k -tiles such that $\mu(A \Delta B) < \epsilon$.

By regularity of μ there exists a compact set $K \subset A$ and an open set $U \subset S^2$ with $K \subset A \subset U$ and $\mu(U \setminus K) < \epsilon$. Since the diameter of tiles goes to 0 uniformly with the order of the tiles, we can choose $k \in \mathbb{N}$ so large that every k -tile that meets K is contained in the open neighborhood U of K . Define

$$A = \bigcup \{X \in \mathbf{X}^k : X \cap K \neq \emptyset\}.$$

Then $K \subset A \subset U$. This implies $A \Delta B \subset U \setminus K$, and so

$$\mu(A \Delta B) \leq \mu(U \setminus K) < \epsilon$$

as desired. The proof is complete. \square

Lemma 20.6. *Let \mathcal{S} be the collection of all sets consisting of the empty set and the interiors of all cells. Then \mathcal{S} is a semi-algebra generating the Borel σ -algebra \mathcal{B} on S^2 .*

Proof. We first show that \mathcal{S} is a semi-algebra by verifying conditions (i)–(iii) of a semi-algebra.

Condition (i): By definition of \mathcal{S} we have $\emptyset \in \mathcal{S}$.

Condition (ii): Let $A, B \in \mathcal{S}$. In order to show that $A \cap B \in \mathcal{S}$, we may assume that $A = \text{int}(\sigma)$ and $B = \text{int}(\tau)$, where σ is a k -cell and τ is an l -cell, and $k \geq l$. Let $p \in \text{int}(\tau)$ be arbitrary. Then by Lemma 4.2 there exists a unique k -cell c with $p \in \text{int}(c)$. Since \mathcal{D}^k is a refinement of \mathcal{D}^l , there exists a unique l -cell τ' with $\text{int}(c) \subset \text{int}(\tau')$ (see Lemma 4.7). Then τ and τ' are both l -cells containing the point p in their interiors. This implies that $\tau' = \tau$, and so $\text{int}(c) \subset \text{int}(\tau)$.

It follows that $\text{int}(\tau)$ can be written as a disjoint union of interiors of k -cells. This implies that either $A \cap B = \text{int}(\sigma)$ or $A \cap B = \emptyset$. In both cases, $A \cap B \in \mathcal{S}$.

Condition (iii): Let $A \in \mathcal{S}$ be arbitrary. If $A = \emptyset$, then $S^2 \setminus A = S^2$, and so $S^2 \setminus A$ is equal to the disjoint union of the interiors of the 0-cells, and so a disjoint union of elements in \mathcal{S} .

If $A = \text{int}(\tau)$ where τ is a k -cell, then $S^2 \setminus A$ is the disjoint union of the interiors of all k -cells distinct from τ . Again $S^2 \setminus A$ is a disjoint union of sets in \mathcal{S} .

So \mathcal{S} is indeed a semi-algebra.

\mathcal{S} generates \mathcal{B} : Let \mathcal{A} be the smallest σ -algebra containing \mathcal{S} . Since \mathcal{S} consists of Borel sets, we have $\mathcal{A} \subset \mathcal{B}$. So in order to show that $\mathcal{A} = \mathcal{B}$ it suffices to establish that $U \in \mathcal{A}$ for each open subset U of S^2 .

Let $p \in U$ be arbitrary. Then for each $k \in \mathbb{N}_0$ the point p is contained in the interior of some k -cell. Since F is expanding, the diameter of k -cells approaches 0 as $k \rightarrow \infty$. Hence there exists a cell c with $p \in \text{int}(c) \subset U$. This implies that U is a union of elements in \mathcal{S} . Since for each $k \in \mathbb{N}_0$ there are only finitely many k -cells, the collection \mathcal{S} is countable, and so U is a countable union of elements in \mathcal{S} . Hence $U \in \mathcal{A}$ as desired. \square

Proposition 20.7. *There exists a unique probability measure μ_F on S^2 such that for each $X \in \mathbf{X}^k$, $k \in \mathbb{N}_0$, we have*

$$(20.11) \quad \mu_F(X) = \begin{cases} w \deg(F)^{-k}, & \text{if } X \text{ is white,} \\ b \deg(F)^{-k} & \text{if } X \text{ is black.} \end{cases}$$

We have $\mu_F(E^\infty) = 0$. Moreover, the measure μ_F is F -invariant, F is mixing for μ_F , and $h_{\mu_F}(F) = \log(\deg(F))$.

So in particular, edges and vertices are sets of μ_F -measure zero, and F is ergodic for μ_F .

Proof. The proof proceeds in several steps.

Construction of μ_F and $\mu_F(E^\infty) = 0$: For each k -tile X , $k \in \mathbb{N}_0$, put $w(X) = w(\deg F)^{-k}$ if X is white and $w(X) = b(\deg F)^{-k}$ if X is black.

If $X \in \mathbf{X}^k$ is white, then w_1 is the number of white $(k+1)$ -tiles contained in X , and b_1 the number of black $(k+1)$ -tiles contained in X . Since $w_1 + w'_1 = \deg(F)$, we have

$$\begin{aligned} \sum_{Y \in \mathbf{X}^{k+1}, Y \subset X} w(Y) &= \frac{w_1 w + b_1 b}{\deg(F)^{k+1}} = \frac{w_1 b_1 + b_1 w'_1}{(b_1 + w'_1) \deg(F)^{k+1}} \\ &= \frac{b_1}{(b_1 + w'_1) \deg(F)^k} = w(X). \end{aligned}$$

A similar equation is also true for black k -tiles. If we iterate these identities, we get

$$(20.12) \quad \sum_{Y \in \mathbf{X}^{k+m}, Y \subset X} w(Y) = w(X)$$

for all $k, m \in \mathbb{N}_0$ and all $X \in \mathbf{X}^k$.

For $A \subset S^2$ we now define

$$(20.13) \quad \mu^*(A) = \inf_{\mathcal{U}} \sum_{X \in \mathcal{U}} w(X),$$

where the infimum is taken over all covers \mathcal{U} of A by tiles (not necessarily of the same order). By subdividing tiles into tiles of high order and using (20.12), one sees that in the infimum in the definition of $\mu^*(A)$ it is enough to only consider covers by tiles whose orders exceed a given number k . Based on this and the fact that $\max_{x \in \mathbf{X}^k} \text{diam}(X) \rightarrow 0$ as $k \rightarrow \infty$, it is clear that μ^* is a *metric outer measure*, i.e., if $A, B \subset S^2$ are sets with $\text{dist}(A, B) > 0$ (with respect to some base metric on S^2), then

$$\mu^*(A \cup B) = \mu^*(A) + \mu^*(B).$$

It is a known fact that the restriction of a metric outer measure to the σ -algebra of Borel sets is a measure. We denote this restriction of μ^* by μ_F .

If $A \subset S^2$ is compact, then it is enough to consider only finite covers by tiles in (20.13). Indeed, suppose $\mathcal{U} = \{X_i : i \in \mathbb{N}\}$ is an infinite cover of the compact set $A \subset S^2$ by tiles, and let $\epsilon > 0$ be arbitrary. It follows from Lemma 20.2 that for each $i \in \mathbb{N}$ we can find a finite collection \mathcal{U}_i of tiles such that $X_i \subset \text{int}(X'_i)$, where

$$X'_i = \bigcup_{X \in \mathcal{U}_i} X$$

and

$$\sum_{X \in \mathcal{U}_i} w(X) \leq w(X_i) + \epsilon/2^i.$$

Finitely many sets $\text{int}(X'_{i_1}), \dots, \text{int}(X'_{i_m})$ will cover A . Then

$$\mathcal{U}' = \mathcal{U}_{i_1} \cup \dots \cup \mathcal{U}_{i_m}$$

is a finite collection of tiles that covers A , and we have

$$\sum_{X \in \mathcal{U}'} w(X) \leq \sum_{X \in \mathcal{U}} w(X) + \epsilon.$$

Since $\epsilon > 0$ was arbitrary, we conclude that for compact sets A we get the same infimum in (20.13) if we restrict ourselves to finite covers by tiles.

A consequence of this is that $\mu_F(X) = w(X)$ for each $X \in \mathbf{X}^k$, $k \in \mathbb{N}_0$. Indeed, by definition of μ_F we obviously have $\mu_F(X) \leq w(X)$.

For an inequality in the opposite direction, it is enough to consider an arbitrary finite cover \mathcal{U} of X by tiles. By subdividing the tiles in \mathcal{U} if necessary, we may also assume that they all have the same order l and that $l \geq k$. Since \mathcal{U} is a cover of X and l -tiles have pairwise disjoint interiors, this implies that $Y \in \mathcal{U}$ whenever $Y \in \mathbf{X}^l$ and $Y \subset X$. Hence

$$w(X) = \sum_{Y \in \mathbf{X}^l, Y \subset X} w(Y) \leq \sum_{Y \in \mathcal{U}} w(Y).$$

Taking the infimum over all \mathcal{U} we get $w(X) \leq \mu_F(X)$ as desired.

Since $\mu_F(X) = w(X)$ for all tiles X , we have (20.11).

It follows from Lemma 20.2 and the definition of μ_F , that if e is an edge, then $\mu_F(e) = 0$. Since E^∞ is the (countable) union of all edges, we have $\mu_F(E^\infty) = 0$. This also shows that

$$\mu_F(S^2) = \sum_{X \in \mathbf{X}^0} \mu_F(X) = \sum_{X \in \mathbf{X}^0} w(X) = w + b = 1,$$

and so μ_F is a probability measure.

Uniqueness of μ_F : Suppose that μ is another probability measure on S^2 satisfying the analog of (20.11). Then from Lemma 20.2 it follows that each edge is a set of μ -measure zero. Hence $\mu_F(\text{int}(c)) = \mu(\text{int}(c))$ for all cells c . Since the empty set together with the interiors of all cells form a semi-algebra \mathcal{S} generating the Borel σ -algebra on S^2 (see Lemma 20.6), we conclude that $\mu = \mu_F$.

μ_F is F -invariant: To show that μ_F is F -invariant, it is enough to verify that

$$(20.14) \quad \mu_F(F^{-1}(A)) = \mu_F(A)$$

for all sets A in the semi-algebra \mathcal{S} . This is true if $A = \emptyset$.

Edges are sets of measure zero, and the preimage of an edge is a finite union of edges (see Proposition 6.1 (ii)). This implies that (20.14) holds if $A = \text{int}(e)$ for an edge e , or if $A = \{v\}$ for a vertex v (since every vertex is contained in an edge).

Moreover, if X is a tile, then $X \setminus \text{int}(X)$ is a union of edges, and so we have $\mu(F^{-1}(\text{int}(X))) = \mu_F(F^{-1}(X))$ and $\mu_F(\text{int}(X)) = \mu_F(X)$. So in order to establish (20.14) it remains to show that

$$\mu_F(F^{-1}(X)) = \mu_F(X)$$

for all tiles X . To see this note that if X is a k -tile, then $F^{-1}(X)$ is a union of $\deg(F)$ $(k+1)$ -tiles that have the same color as X . Since the intersection of any two distinct $(k+1)$ -tiles is contained in a union of edges, and hence a set of measure zero, it follows that from (20.11) that

$$\mu_F(F^{-1}(X)) = \deg(F) \frac{\mu_F(X)}{\deg(F)} = \mu_F(X).$$

F is mixing for μ_F : It suffices to show that for all sets A and B in the semi-algebra \mathcal{S} we have

$$\mu_F(F^{-m}(A) \cap B) \rightarrow \mu_F(A)\mu_F(B)$$

as $m \rightarrow \infty$. Based on the fact that edges are sets of measure zero and that the preimage of each edge under F is a finite union of edges, for this it suffices to show that for all tiles X and Y we have

$$\mu_F(F^{-m}(X) \cap Y) \rightarrow \mu_F(X)\mu_F(Y)$$

as $m \rightarrow \infty$.

So let $k, l, m \in \mathbb{N}_0$, $X = X^k \in \mathbf{X}^k$ and $Y = Y^l \in \mathbf{X}^l$ be arbitrary. We may assume that $m \geq l$. Then $F^{-m}(X^k)$ is a union of $(m+k)$ -tiles that have the same color as X^k . Since edges are sets of measure zero

and the $(m+k)$ -tiles subdivide the l -tile Y^l , it follows that

$$(20.15) \quad \mu_F(F^{-m}(X^k) \cap Y^l) = \frac{\mu_F(X^k)}{\deg(F)^m} \cdot \#\{Z^{m+k} \in \mathbf{X}^{m+k} : Z^{m+k} \subset Y^l, F^m(Z^{m+k}) = X^k\}.$$

Let $X^0 \in \mathbf{X}^0$ be the unique tile with $X^k \subset X^0$, and $Y^0 := F^l(Y^l) \in \mathbf{X}^0$. We assume that X^0 and Y^0 are both white; the other cases are similar.

Every $(m+k)$ -tile Z^{m+k} lies in a unique “parent” m -tile Z^m . Since Y^l is an l -tile and $m \geq l$, we have $Z^{m+k} \subset Y^l$ if and only if $Z^m \subset Y^l$. If $F^m(Z^{m+k}) = X^k$, then $F^m(Z^m)$ is a 0-tile containing X^k , and so $F^m(Z^m) = X^0$. Conversely, if Z^m is an m -tile and $F^m(Z^m) = X^0$, then it follows from Lemma 18.4 (i) that $Z^{m+k} := (F^m|Z^m)^{-1}(X^k)$ is the unique $(m+k)$ -tile with $Z^{m+k} \subset Z^m$ and $F^m(Z^{m+k}) = X^k$. These statements imply that the map $Z^{m+k} \mapsto Z^m$ that assigns to each $(m+k)$ -tile Z^{m+k} its unique parent m -tile Z^m induces a bijection between the sets $\{Z \in \mathbf{X}^{m+k} : Z^{m+k} \subset Y^l, F^m(Z^{m+k}) = X^k\}$ and $\{Z^m \in \mathbf{X}^m : Z^m \subset Y^l, F^m(Z^m) = X^0\}$. Hence

$$(20.16) \quad \#\{Z^{m+k} \in \mathbf{X}^{m+k} : Z^{m+k} \subset Y^l, F^m(Z^{m+k}) = X^k\} = \#\{Z^m \in \mathbf{X}^m : Z^m \subset Y^l, F^m(Z^m) = X^0\}.$$

Since X^0 is white, the last quantity is equal to the number of white m -tiles contained in Y^l . Applying the homeomorphism $F^l|Y^l$, we see that this number is equal to w_{m-l} , the number of white $(m-l)$ -tiles contained in the white 0-tile $Y^0 = F^l(Y^l)$. Combing this with (20.15), (20.16), and Lemma 20.1, we get

$$\mu_F(F^{-m}(X) \cap Y) = \frac{\mu_F(X)}{\deg(F)^m} \cdot w_{m-l} \rightarrow \frac{\mu_F(X)}{\deg(F)^l} \cdot w = \mu_F(X) \mu_F(Y)$$

as $m \rightarrow \infty$.

The identity $h_{\mu_F}(F) = \log(\deg(F))$: According to Lemma 20.5 the measurable partition $\xi = \mathbf{X}^1$ is a generator for (F, μ_F) , and so $h_{\mu_F}(F) = h_{\mu_F}(F, \xi)$. Moreover, for each $k \in \mathbb{N}$, the measurable partition ξ_F^k is equivalent to the measurable partition \mathbf{X}^k given by the k -tiles. Hence

$$\begin{aligned} H_{\mu_F}(F, \xi_F^k) &= H_{\mu_F}(F, \mathbf{X}^k) = \sum_{X \in \mathbf{X}^k} \mu_F(X) \log(1/\mu_F(X)) \\ &= w \log(\deg(F)^k/w) + b \log(\deg(F)^k/b) \\ &= k \log(\deg(F)) + w \log(1/w) + b \log(1/b). \end{aligned}$$

This implies

$$h_{\mu_F}(F) = h_{\mu_F}(F, \xi) = \lim_{k \rightarrow \infty} \frac{1}{k} H_{\mu_F}(F, \xi_F^k) = \log(\deg(F)).$$

The proof is complete. \square

Corollary 20.8. $h_{top}(f) = \log(\deg(f))$.

Proof. We know that $h_{top}(F) \leq \log(\deg(F))$, as we have seen in the proof of Lemma 20.4. Moreover, we have $h_{\mu_F}(F) = \log(\deg(F))$ by Proposition 20.7, and so $h_{top}(F) \geq \log(\deg(F))$ by the variational principle. It follows that $h_{top}(F) = \log(\deg(F))$. Since $F = f^n$ and so $\deg(F) = \deg(f)^n$ and $h_{top}(F) = nh_{top}(f)$, the claim follows. \square

Theorem 20.9. *The measure μ_F is the unique measure of maximal entropy for f , i.e., the unique f -invariant probability measure μ_F with $h_{\mu_F}(f) = h_{top}(f)$. Moreover, f is mixing for μ_F .*

In particular, f is ergodic for μ_F .

Proof. We first show uniqueness. So let ν be a probability measure that is f -invariant and satisfies $h_\nu(f) = h_{top}(f)$. Then ν is F -invariant and satisfies

$$(20.17) \quad h_\nu(F) = nh_\nu(f) = nh_{top}(f) = h_{top}(F) = \log(\deg(F)).$$

We will show that this implies that $\nu = \mu_F$. The proof proceeds in several steps.

We can write ν as a convex combination $\nu = \beta\nu_s + (1 - \beta)\nu_a$, where ν_s is a probability measure that is singular with respect to μ_F and ν_a is a probability measure that is absolutely continuous with respect to μ_F . Since μ_F and ν are F -invariant, it follows from the uniqueness of the decomposition of a measure into singular and absolutely continuous part that ν_s and ν_a are also F -invariant. Since F is ergodic for μ_F and ν_a is F -invariant and absolutely continuous with respect to μ_F , it follows that $\nu_a = \mu_F$.

If $\beta = 0$, then $\nu = \mu_F$ and we are done. If $\beta > 0$, then we can use the equation

$$\begin{aligned} \log(\deg(F)) &= h_\nu(F) = \beta h_{\nu_s}(F) + (1 - \beta) h_{\mu_F}(F) \\ &= \beta h_{\nu_s}(F) + (1 - \beta) \log(\deg(F)), \end{aligned}$$

to conclude that

$$h_{\nu_s}(F) = \log(\deg(F)).$$

We will show that this is impossible by proving that for every F -invariant probability measure ν that is singular with respect to μ_F we must have

$$h_\nu(F) < \log(\deg(F)).$$

The uniqueness of μ_F will then follow.

So let ν be such a measure and consider the union E^∞ of all edges. Assume first that $\nu(E^\infty) > 0$. By (20.9) we can then write ν as a convex combination $\nu = \alpha\nu_1 + (1 - \alpha)\nu_2$ of two F -invariant probability measures ν_1 and ν_2 , where $\alpha = \nu(E^\infty)$, ν_1 is concentrated on E^∞ , and ν_2 on $S^2 \setminus E^\infty$.

Since ν_1 is F -invariant, we have $\nu_1(F^{-k}(\mathcal{C})) = \nu_1(\mathcal{C})$ for all $k \in \mathbb{N}_0$. On the other hand, $\mathcal{C} \subset F^{-k}(\mathcal{C})$, and so $\nu_1(F^{-k}(\mathcal{C}) \setminus \mathcal{C}) = 0$. This implies that $\nu_1(E^\infty \setminus \mathcal{C}) = 0$. So ν_1 is actually concentrated on \mathcal{C} . Therefore, by the variational principle and by Lemma 20.4 we have

$$h_{\nu_1}(F) = h_{\nu_1}(F|\mathcal{C}) \leq h_{top}(F|\mathcal{C}) < \log(\deg(F)).$$

We also have $h_{\nu_2}(F) \leq h_{top}(F) = \log(\deg(F))$, and so

$$h_\nu(F) = \alpha h_{\nu_1}(F) + (1 - \alpha)h_{\nu_2}(F) < \log(\deg(F)).$$

In this case we are done.

In the remaining case we have $\nu(E^\infty) = 0$. Then by Lemma 20.5 $\xi = \mathbf{X}^1$ is a generator for (F, ν) , and so $h_\nu(F) = h_\nu(F, \nu)$. In particular,

$$h_\nu(F) = \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{X \in \mathbf{X}^k} \nu(X) \log(1/\nu(X)),$$

and the limit is given by the infimum of the sequence elements.

Since ν and μ_F are mutually singular, we can find a Borel set $A \subset S^2$ with $\nu(A) = 1$ and $\mu_F(A) = 0$. Using inner regularity of ν and outer regularity of μ_F , for each $\epsilon > 0$ we can find a compact set $K \subset S^2$ and an open set $U \subset S^2$ with $K \subset A \subset U$, $\nu(K) > 1 - \epsilon$ and $\mu_F(U) < \epsilon$. If k is sufficiently large, then we can cover the set K by k -tiles contained in U .

Using this for smaller and smaller $\epsilon > 0$, we conclude that for each $k \in \mathbb{N}$ we can find a set $M_k \subset \mathbf{X}^k$ such that for $A_k := \bigcup\{X \in M_k\}$ we have $\nu(A_k) \rightarrow 1$ and $\mu_F(A_k) \rightarrow 0$ as $k \rightarrow \infty$. Note that $\mu_F(X) \geq c \deg(F)^{-k}$ for each $X \in \mathbf{X}^k$, where $c > 0$ is independent of k and X . Hence

$$\#M_k \leq \frac{1}{c} \mu_F(A_k) \deg(F)^k.$$

We also have $\#\mathbf{X}^k = 2 \deg(F)^k$.

Note that if $M \subset \mathbf{X}^k$ is arbitrary and $A = \bigcup_{X \in M} X$, then by concavity of the function $x \mapsto \phi(x) = x \log(1/x)$, we have

$$\begin{aligned} \sum_{X \in M} \nu(X) \log(1/\nu(X)) &\leq \#M \cdot \phi\left(\frac{1}{\#M} \sum_{X \in M} \nu(X)\right) \\ &= \nu(A) \log(\#M/\nu(A)) \leq \nu(A) \log(\#M) + 1/e. \end{aligned}$$

Hence

$$\begin{aligned}
k \log(\deg(F)) &= kh_\nu(F) \\
&\leq \sum_{X \in \mathbf{X}^k} \nu(X) \log(1/\nu(X)) \\
&= \sum_{X \in M_k} \dots + \sum_{X \in \mathbf{X}^k \setminus M_k} \dots \\
&\leq \nu(A_k) \log(\#M_k) + \nu(S^2 \setminus A_k) \log(\#\mathbf{X}^k) + C_1 \\
&\leq \nu(A_k) \log(\mu_F(A_k)) + (\nu(A_k) + \nu(S^2 \setminus A_k)) \log(\deg(F)^k) + C_2 \\
&= \nu(A_k) \log(\mu_F(A_k)) + k \log(\deg(F)) + C_2.
\end{aligned}$$

Here the constants C_1 and C_2 do not depend on k . An inequality of this type is impossible as

$$\nu(A_k) \log(\mu_F(A_k)) \rightarrow -\infty$$

for $k \rightarrow \infty$.

This shows that if there is a measure of maximal entropy for f , then it has to agree with μ_F . Actually, we proved that μ_F is the unique measure of maximal entropy for F .

We now show that μ_F is f -invariant and a measure of maximal entropy for f . Indeed, the measure $f_*\mu_F$ is F -invariant and the triple $(S^2, F, f_*\mu)$ is a factor of (S^2, F, μ_F) by the map f . It follows that $h_{f_*\mu}(F) \leq h_\mu(F)$. Iterating this and noting that $f_*^n\mu_F = F_*\mu_F = \mu_F$ by F -invariance of μ_F , leads to

$$h_{\mu_F}(F) = h_{f_*^n\mu_F}(F) \leq h_{f_*^{n-1}\mu_F}(F) \leq \dots \leq h_{f_*\mu_F}(F) \leq h_{\mu_F}(F).$$

Hence $h_{f_*\mu_F}(F) = h_{\mu_F}(F)$, and so $f_*\mu_F$ is a measure of maximal entropy for F . By uniqueness of μ_F we have $f_*\mu_F = \mu_F$ showing that μ_F is f -invariant. From the computation (20.17) we see that $h_{\mu_F}(f) = \log(\deg(f)) = h_{top}(f)$. So μ_F is a measure of maximal entropy for f , and by the first part of the proof we know that it is the unique such measure.

It remains to show that f is mixing for μ_F . Indeed, since F is mixing for μ_F (Proposition 20.7) and μ_F is f -invariant, we have that for all $m \in \{0, \dots, n-1\}$, and all Borel sets $A, B \subset S^2$,

$$\begin{aligned}
\mu_F(f^{-(nl+m)}(A) \cap B) &= \mu_F(F^{-l}(f^{-m}(A)) \cap B) \rightarrow \\
&\mu_F(f^{-m}(A))\mu_F(B) = \mu_F(A)\mu_F(B)
\end{aligned}$$

as $l \rightarrow \infty$. Hence we get the desired relation

$$\mu_F(f^{-k}(A) \cap B) \rightarrow \mu_F(A)\mu_F(B)$$

as $k \rightarrow \infty$. The proof is complete. \square

A metric space X is called (Ahlfors) Q -regular, where $Q > 0$, if there exists a (Borel) measure μ such that

$$\frac{1}{C}R^Q \leq \mu(\overline{B}(x, R)) \leq CR^Q$$

for all closed balls $\overline{B}(x, R)$ with $x \in X$ and $0 < R \leq \text{diam}(X)$, where $C \geq 1$ is independent of the ball.

Proposition 20.10. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map without periodic critical points, and d be a visual metric with expansion factor Λ . Then (S^2, d) is Ahlfors Q -regular with*

$$Q = \frac{\log(\deg(f))}{\log(\Lambda)}.$$

Proof. Consider the iterate $F = f^n$ chosen as in the beginning of this section, and let d be a visual metric for f with expansion factor Λ . Then d is a visual metric for F with expansion factor $\Lambda' = \Lambda^n$.

Consider an arbitrary closed ball $\overline{B}(x, R)$, where $x \in S^2$ and $0 < R \leq \text{diam}(S^2)$. We use the sets $U^m(x)$ as defined in (8.11) for the map F . Since F does not have periodic critical points, the length of the cycle of each vertex is uniformly bounded (see Lemma 17.1 and Lemma 7.2 (i)). This implies that each set $U^m(x)$ consists of a uniformly bounded number of tiles. In particular, these sets are closed. Moreover, by Lemma 8.12 (applied to F) we have

$$U^{m+n_0}(x) \subset \overline{B}(x, R) \subset U^{m-n_0}(x),$$

where $m = \lceil -\log(R)/\log(\Lambda') \rceil$ and $n_0 \in \mathbb{N}_0$ is a constant independent of the ball.

Noting that

$$Q = \frac{\log(\deg(F))}{\log(\Lambda')}$$

and using (20.11), we get

$$\begin{aligned} \mu_F(U^{m+n_0}(x)) &\asymp \mu_F(U^{m-n_0}(x)) \asymp \deg(F)^{-m} \\ &\asymp \exp(\log(R) \log(\deg(F))/\log(\Lambda')) = R^Q, \end{aligned}$$

and so $\mu_F(\overline{B}(x, R)) \asymp R^Q$. Here the constants $C(\asymp)$ are independent of the ball. The statement follows. \square

21. CONCLUDING REMARKS AND OPEN PROBLEMS

The authors hope that the results in the present paper will serve as a foundation for future work in the area. Our line of investigation has been continued in [Me09b, Me09c, Yi]. In the following we will discuss a selection of questions and open problems that are worth pursuing.

The basis of our combinatorial approach is the existence of cellular Markov partitions for Thurston maps. We know by Corollary 1.5 that such a cellular Markov partition (related to a two-tile subdivision rule) exists for every sufficiently high iterate of an expanding Thurston map. This leads to the following question:

Problem 1. *Does every expanding Thurston map $f: S^2 \rightarrow S^2$ have a cellular Markov partition?*

In the introduction we conjectured that the answer should be affirmative. To prove the conjecture, one essentially has to construct a connected graph $G \subset S^2$ with $\text{post}(f) \subset G$ that is f -invariant, i.e., $f(G) \subset G$. A different way to phrase the problem is to ask whether f (and not some iterate f^n) can be described by a (suitably defined) k -tile subdivision rule. Here k would be the number of components of $S^2 \setminus G$.

As we have seen, the dynamics of an expanding Thurston map f generates a class of visual metrics, and so a fractal geometry on the 2-sphere on which it acts. This does not rule out that the map f can actually be described by a smooth model.

Problem 2. *Is every expanding Thurston map $f: S^2 \rightarrow S^2$ topologically conjugate to a smooth expanding Thurston map $g: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$?*

This question was raised by K. Pilgrim. We expect that this is true for at least every sufficiently high iterate of f .

In our examples of rational Thurston maps, the coefficients were always algebraic numbers. In general one can ask:

Problem 3. *Is every expanding rational Thurston map topologically conjugate to a rational Thurston map with algebraic coefficients?*

J. Souto posed this question after a talk by the first author. We expect a positive answer.

We know that every expanding Thurston map f has iterates f^n with invariant curves as in Theorem 1.2. This naturally leads to the question whether one can bound the order of this iterate in terms of some natural invariants of the map.

Problem 4. *Let $f: S^2 \rightarrow S^2$ be an expanding Thurston map. Is there a number $N_0 \in \mathbb{N}$, depending on some natural data such as $\deg(f)$, $\#\text{post}(f)$, and $\Lambda_0(f)$ such that for all $n \geq N_0$ there exists an f^n -invariant Jordan curve $\mathcal{C} \subset S^2$ with $\mathcal{C} \subset \text{post}(f)$?*

Recall that $\Lambda_0(f)$ denotes the combinatorial expansion factor of f (see Section 18).

According to one of our main results every expanding Thurston map has an iterate that can be described by a two-tile subdivision rule (Corollary 1.5). If a map realizes a subdivision rule, it should be possible to extract all relevant information about the map from the combinatorial data. This underlying principle is the common thread for the next three problems.

Every two-tile subdivision rule is realized by a Thurston map that is unique up to Thurston equivalence. In contrast, a Thurston map may realize combinatorially different two-tile subdivision rules. This leads to the following question:

Problem 5. *Suppose two Thurston maps f and g realize different two-tile subdivision rules. How can one decide from combinatorial data whether the maps are Thurston equivalent?*

Of course, there are several simple necessary conditions, such as $\deg(f) = \deg(g)$ and $\#\text{post}(f) = \#\text{post}(g)$, whose validity can easily be checked from the subdivision rules. In addition, the maps have to have the same *critical portrait* (see Example 15.5). A related question, namely when polynomials with the same critical portrait are Thurston equivalent, is answered in [BN].

Problem 6. *Let f be an expanding Thurston map that realizes a two-tile subdivision rule. Is there an effective way to compute the combinatorial expansion factor $\Lambda_0(f)$ from the combinatorial description?*

Since $\Lambda_0(f)$ is defined as a limit, a priori one cannot expect to find $\Lambda_0(f)$ by a finite procedure. However, if f is a Lattès or Lattès-type map (see Example 18.5), then $\Lambda_0(f)$ is the smallest absolute value of any eigenvalue of the matrix describing the underlying torus endomorphism [Yi]. In general, one may speculate that if f realizes a two-tile subdivision rule with underlying cell decompositions \mathcal{D} and \mathcal{D}' , then $\Lambda_0(f)$ is related to the spectral radius of a matrix that is obtained from the incidence relations of the cells in \mathcal{D}' and their images under f in \mathcal{D} .

Problem 7. *Let f be an expanding Thurston map that realizes a two-tile subdivision rule. Is it possible to decide from the subdivision rule whether f is Thurston equivalent to a rational map?*

As discussed in the introduction, Thurston proved a necessary and sufficient condition for a Thurston map to be equivalent to a rational map [DH]. In principle, it is possible to check this criterion if the map is given by a subdivision rule. In practice, the sufficiency part is hard to verify, since it involves the behavior of the map on infinitely many homotopy classes of curves. So in Problem 7 we really ask for a substantially simpler criterion, preferably based on the verification of a finite condition. As we have seen (Theorem 1.10), the case $\# \text{post}(f) = 3$ is completely understood. The case $\# \text{post}(f) = 4$ should also be rather accessible. For understanding the underlying issues it might be worthwhile to reprove Thurston's result by using a combinatorial approach as in the present paper.

From Theorem 1.9 it follows that for the solution of Problem 7 one may equivalently ask if one can decide from the subdivision rule whether S^2 equipped with a visual metric d is quasimetrically equivalent to the standard 2-sphere (i.e., the unit sphere in \mathbb{R}^3). This is closely related to Cannon's conjecture in Geometric Group Theory which can be reformulated as a similar quasimetric equivalence problem (see [Bo] for more background, and [Ca94] for a different approach). The general problem when a metric 2-sphere (S^2, d) is quasimetrically equivalent to the standard 2-sphere is a hard problem that is not completely understood (for results in this direction, see [BK]).

It is well-known that there are many analogies between Complex Dynamics and the theory of Kleinian groups (informally referred to as *Sullivan's dictionary*). It seems very fruitful to explore this systematically for Thurston maps and translate statements for these maps into corresponding group theoretic analogs. Expanding Thurston maps correspond to Gromov hyperbolic groups whose boundary at infinity is a 2-sphere. This is suggested, for example, by Theorem 1.9, which can be seen as an analog of Cannon's Conjecture, or by Remark 8.8. For Gromov hyperbolic groups whose boundary at infinity is a Sierpinski carpet there is an analog of Cannon's conjecture—the Kapovich-Kleiner conjecture. It predicts that these groups arise from some standard situation in hyperbolic geometry. It would be interesting to formulate an analog of this for Thurston maps in the spirit of Theorem 1.9.

If $f: S^2 \rightarrow S^2$ is an expanding Thurston map without periodic critical points, then S^2 equipped with a visual metric d is Ahlfors regular (Proposition 20.10). The infimal exponent Q such that (S^2, d) is quasimetrically equivalent to an Ahlfors Q -regular space is called the (*Ahlfors regular*) *conformal dimension* of (S^2, d) . This is an important

numerical invariant of the fractal geometry of (S^2, d) . Actually, this conformal dimension only depends on f and not on the choice of the visual metric d , since all visual metrics are quasimetrically equivalent.

Problem 8. *Is it possible to determine the conformal dimension of a 2-sphere equipped with a visual metric of an expanding Thurston map f in terms of dynamical data of f ?*

Bonk-Geyer-Pilgrim [Bo] stated a conjecture expressing this conformal dimension in terms of eigenvalues of certain matrices related to the dynamics of f . One half of this conjecture has recently been proved Haïssinsky-Pilgrim [HP08].

To prove his characterization of rational maps among Thurston maps [DH], Thurston used an iteration procedure on a suitable Teichmüller space. This idea may also be fruitful for studying some problems related to our existence criteria for invariant Jordan curves. Namely, in view of Theorem 1.3 one is led to the general question how to find non-trivial sets $K \subset S^2$ with $\text{post}(f) \subset K$ (and maybe additional geometric features) for which there exists an isotopy rel. $\text{post}(f)$ that deforms the set into a subset of its preimage $f^{-1}(K)$. It may be possible to reformulate this as a fixed point problem under iteration of a pull-back operation on a suitable space.

As we discussed above, Problem 1 is equivalent to finding an f -invariant graph G with suitable properties. If one does not insist on describing the map by a cellular Markov partition, then other invariant sets K may lead to other natural combinatorial descriptions of the map or its iterates.

Actually, in many cases one can show the existence of an f^n -invariant arc $A \subset S^2$ with $\text{post}(f) \subset A$. Slightly more general, one can ask if an invariant tree $T \subset S^2$ with $\text{post}(f) \subset T$ exists, i.e., a graph $T \subset S^2$ (in the sense of Section 14) such that $S^2 \setminus T$ is connected.

Problem 9. *Let $f: S^2 \rightarrow S^2$ be expanding Thurston map. Does there exist a tree $T \subset S^2$ with $\text{post}(f) \subset T$ that is invariant under f (or at least under a sufficiently high iterate of f)?*

If such an invariant tree T exists, then $\Omega^0 = S^2 \setminus T$ is a simply connected region that is subdivided by the complementary components Ω^1 of $f^{-1}(T)$. Moreover, on each component Ω^1 the map $f|_{\Omega^1}$ is a homeomorphism onto Ω^0 . A similar statement then also holds for the iterates of f . Since there is only one open “0-tile” Ω^0 , the dynamics of f is described by a *one-tile subdivision rule*. This would simplify the combinatorial description of f ; on the other hand, open tiles then are

only simply connected regions and not necessarily open Jordan regions. So the simplified combinatorial description comes at the price of a more complicated geometry of tiles.

For postcritically finite polynomials such an invariant tree always exists, namely the *Hubbard tree* introduced by Douady and Hubbard, see [DH84]. In fact the Hubbard tree determines the polynomial uniquely up to conjugation by a map $z \mapsto az + b$, $a \in \mathbb{C} \setminus \{0\}, b \in \mathbb{C}$, see [Poi].

Cannon-Floyd-Parry show in [CFP10, Theorem 3.1] that a sufficiently high iterate of each Lattès map has an invariant tree. Note however that “Lattès map” is used there in a more restrictive sense, i.e., it denotes a quotient of a *double* cover of a conformal torus endomorphism.

If one further relaxes the geometric requirements on “tiles”, then other combinatorial descriptions of (expanding) Thurston maps exist. The result presented in Section 9 can be reinterpreted in this vein. This is related to the description of f via “limit spaces” of its *iterated monodromy group* (see [Ne] and the discussion below).

Yet another point of view to obtain combinatorial descriptions of Thurston maps is to consider sets that are “dual” to invariant sets containing the postcritical set. To illustrate this, we consider the Lattès map g from Section 1.2, and again represent the sphere S^2 as a pillow as in Figure 1. Let K be the subset of the pillow that is of the union of four Jordan curves, namely the two curves consisting of the points at distance $1/3$ and at $2/3$, respectively, from the bottom edge of the pillow, and the two closed curves of the points at distance $1/3$ and $2/3$ from the left edge. Then K is g -invariant and for each $n \in \mathbb{N}_0$ the set $g^{-n}(K)$ is a graph that is the 1-skeleton of a cell decomposition of the pillow. Moreover, if \mathcal{C} is the (g -invariant) boundary of the pillow, then $g^{-n}(K)$ is a dual graph to $g^{-n}(\mathcal{C})$.

In this paper we mostly considered expanding Thurston maps. One may ask whether combinatorial descriptions exist for more general types of maps.

Problem 10. *Let $f: S^2 \rightarrow S^2$ be a Thurston map (not necessarily expanding). Is there a Jordan curve $\mathcal{C} \subset S^2$ with $\text{post}(f) \subset \mathcal{C}$ that is invariant for some iterate f^n ? Are there other natural decompositions of the sphere S^2 that are invariant (in a suitable sense) under f or some iterate f^n ?*

When f is a polynomial natural decompositions are obtained via *external rays* (see [DH84]). Closely related are *Yoccoz puzzles* (see [Hu] and [Mi00]). A lack of such a description for rational maps is

one of the main reasons why their study is harder than the study of polynomials.

Similarly, one can ask if and how our results extend to maps that are not postcritically-finite. The theory of *coarse expanding dynamical systems* developed in [HP09] should be relevant here.

Problem 11. *Let $R: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ be a rational map (not necessarily post-critically-finite) whose Julia set is the whole Riemann sphere $\widehat{\mathbb{C}}$. Does there exist a natural combinatorial description of R or some iterate R^n ?*

The rational maps $R: \widehat{\mathbb{C}} \rightarrow \widehat{\mathbb{C}}$ of a given degree $d \geq 2$ form a complex manifold \mathcal{R}_d of dimension $2d + 1$. M. Rees has shown [Re] that the set of points in \mathcal{R}_d where the corresponding rational map R has as Julia set equal to the whole sphere has positive measure with respect to the natural measure class on \mathcal{R}_d . These points/rational maps are obtained by a slight perturbation of a certain expanding Thurston map $R_0 \in \mathcal{R}_d$. It would be interesting to find combinatorial descriptions of expanding Thurston maps that change under such perturbations in a controlled manner.

Our combinatorial approach seems quite relevant in the study of the *iterated monodromy group* of a Thurston map $f: S^2 \rightarrow S^2$. To state the basic definitions, let $G = \pi_1(S^2 \setminus \text{post}(f), p)$ be the fundamental group of $S^2 \setminus \text{post}(f)$ represented by homotopy classes of loops in $S^2 \setminus \text{post}(f)$ with initial point $p \in S^2 \setminus \text{post}(f)$. Then there exists a natural action of G on the disjoint union T of the sets $f^{-n}(p)$, $n \in \mathbb{N}_0$. Namely, if $x \in f^{-n}(p) \subset T$, and $g \in G$ is represented by a loop α starting at p , then we can lift α by the map f^n to a path $\tilde{\alpha}$ with initial point x . If y is the endpoint of $\tilde{\alpha}$, then $y \in f^{-n}(p)$, and we define $g(x) := y$.

The action $G \curvearrowright T$ is not effective in general, i.e., there may be elements $g \in G$ that act as the identity on T . Let H be the ineffective kernel of this action, i.e., the normal subgroup of G consisting of all elements in G acting as the identity on T . Then the *iterated monodromy group* of f is defined as the quotient group $\Gamma = G/H$. The action of G on T induces a natural action $\Gamma \curvearrowright T$. By the same method as in the proof of Theorem 1.6 one can find a coding of the elements in T by words in a finite alphabet. Then it is not hard to write down explicit recursive relations for the action of suitably chosen generators of Γ on these words. In principle, this describes the group Γ completely, but it seems hard in specific cases to get a complete understanding of the structure of the group.

Problem 12. *What is the iterated monodromy group of an expanding Thurston map? In particular, what is the growth behavior of such a group?*

The last question is interesting, because there is an example of a (non-expanding) Thurston map (namely, the polynomial $f(z) = z^2 + i$) whose iterated monodromy group is a group of *intermediate growth* [BP]. Formerly, these groups had been regarded as rather exotic objects.

We conclude our discussion with some remarks on higher dimensions. The concept of postcritically-finite maps has been generalized to several complex variables and the dynamics of these maps has been studied (see, for example, [FS92, FS94, Jo]). A possible different direction is the theory of quasiregular mappings on $\widehat{\mathbb{R}}^n = \mathbb{R}^n \cup \{\infty\}$.

Problem 13. *Develop a theory of quasiregular maps in higher dimensions that generalizes the theory of Thurston maps in dimension 2.*

It is not even clear what the basic definitions should be. It seems reasonable to require that a *quasiregular Thurston map* $f: \widehat{\mathbb{R}}^n \rightarrow \widehat{\mathbb{R}}^n$ is *uniformly quasiregular* (i.e., there exists $K \geq 1$ such that f^n is K -quasiregular for each $n \in \mathbb{N}$) and that the postcritical set $\text{post}(f)$ of f (i.e., the forward orbit of the branch set of f) is “small” and “non-exotic”. For example, one could require that $\text{post}(f)$ is an embedded cell complex in $\widehat{\mathbb{R}}^n$ of codimension 2.

REFERENCES

- [AH] P.S. Alexandroff and H. Hopf *Topologie I*, Springer, Berlin, 1935.
- [BN] L. Bartholdi and V. Nekrashevych, *Thurston equivalence of topological polynomials*, Acta Math. 197 (2006), 1–51.
- [Bo] M. Bonk, *Quasiconformal geometry of fractals*, Proc. Internat. Congr. Math., (Madrid, Spain, 2006), Europ. Math. Soc., Zürich, 2006, pp. 1349–1373.
- [BK] M. Bonk and B. Kleiner, *Quasisymmetric parametrizations of two-dimensional metric spheres*, Invent. Math. 150 (2002), 127–183.
- [BS] P.L. Bowers and K. Stephenson, *A “regular” pentagonal tiling of the plane*, Conform. Geom. Dyn. 1 (1997), 58–86.
- [Bur] R.B. Burckel, *Inessential maps and classical Euclidean topology*, Jahrbuch Überblicke Mathematik 1981, Eds. S.D. Chatterji et al., Bibliographisches Inst., Mannheim, 1981, pp. 119–137.
- [Bu] P. Buser, *Geometry and spectra of compact Riemann surfaces*, Birkhäuser, Boston, MA, 1992.
- [BuS] S. Buyalo and V. Schroeder, *Elements of asymptotic geometry*, Europ. Math. Soc., Zürich, 2007.

- [BP] K.-U. Bux and R. Pérez, *On the growth of iterated monodromy groups*. Topological and asymptotic aspects of group theory, Contemp. Math., 394, Amer. Math. Soc., Providence, RI, 2006, pp. 61–76.
- [Ca94] J.W. Cannon, *The combinatorial Riemann mapping theorem*, Acta Math. 173 (1994), 155–234.
- [Ca] J.W. Cannon, *The recognition problem: what is a topological manifold?*, Bull. Amer. Math. Soc. 84 (1978), 832–866.
- [CFKP] J.W. Cannon, W.J. Floyd, R. Kenyon, and W.R. Parry, *Constructing rational maps from subdivision rules*, Conform. Geom. Dyn. 7 (2003), 76–102.
- [CFP01] J.W. Cannon, W.J. Floyd, and W.R. Parry, *Finite subdivision rules*, Conform. Geom. Dyn. 5 (2001), 153–196.
- [CFP06a] J.W. Cannon, W.J. Floyd, and W.R. Parry, *Expansion complexes for finite subdivision rules. I*. Conform. Geom. Dyn. 10 (2006), 63–99.
- [CFP06b] J.W. Cannon, W.J. Floyd, and W.R. Parry, *Expansion complexes for finite subdivision rules. II*. Conform. Geom. Dyn. 10 (2006), 326–354.
- [CFP07] J.W. Cannon, W.J. Floyd, and W.R. Parry, *Constructing subdivision rules from rational maps*, Conform. Geom. Dyn. 11 (2007), 128–136.
- [CFP10] J.W. Cannon, W.J. Floyd, and W.R. Parry, *Lattès maps and finite subdivision rules*, Conform. Geom. Dyn. 14 (2010), 113–140.
- [CG] L. Carleson and T.H. Gamelin, *Complex dynamics*, Springer, 1993.
- [Da] R.J. Daverman, *Decompositions of manifolds*, Academic Press, Orlando, FL, 1986.
- [DH84] A. Douady and J.H. Hubbard, *Étude dynamique des polynômes complexes. I& II*. Publ. Math. d’Orsay, Univ. de Paris-Sud, Dépt. de Math., Orsay, 1984–85.
- [DH] A. Douady and J.H. Hubbard, *A proof of Thurston’s topological characterization of rational functions*, Acta Math. 171 (1993), 263–297.
- [FS92] J.E. Fornæss and N. Sibony, *Critically finite rational maps on \mathbf{P}^2* . The Madison Symposium on Complex Analysis, Contemp. Math. 137, Amer. Math. Soc., Providence, RI, 1992, pp. 245–260.
- [FS94] J.E. Fornæss and N. Sibony, *Complex dynamics in higher dimension. I*. Complex analytic methods in dynamical systems (Rio de Janeiro, 1992), Astérisque 222 (1994), 201–231.
- [La] S. Lattès, *Sur l’itération des substitutions rationnelles et les fonctions de Poincaré*, C. R. Acad. Sci. Paris 166 (1918), 26–28.
- [GH] E. Ghys & P. de la Harpe, Eds., *Sur les groupes hyperboliques d’après Mikhael Gromov*, Progress in Math. 83, Birkhäuser, Boston, 1990.
- [Ha] A. Hatcher, *Algebraic topology*, Cambridge Univ. Press, Cambridge, 2002.
- [HP08] P. Haïssinsky and K. Pilgrim, *Thurston obstructions and Ahlfors regular conformal dimension*, J. Math. Pures Appl. (9) 90 (2008), 229–241.
- [HP09] P. Haïssinsky and K. Pilgrim, *Coarse expanding conformal dynamics*, Astérisque 325 (2009).
- [He] J. Heinonen, *Lectures on analysis on metric spaces*, Springer, New York, 2001.
- [Hu] J. H. Hubbard, *Local connectivity of Julia sets and bifurcation loci: three theorems of J.-C. Yoccoz*. Topological methods in modern mathematics

- (Stony Brook, NY, 1991), Publish or Perish, Houston, TX, 1993, pp. 467–511,
- [IM] T. Iwaniec and G. Martin, *Geometric function theory and non-linear analysis*, Clarendon Press, Oxford, 2001.
 - [Ja] M.V. Jakobson, *Markov partitions for rational endomorphisms of the Riemann sphere*, Multicomponent random systems, Adv. Probab. Related Topics 6, Dekker, New York, 1980, pp. 381–396.
 - [Jo] M. Jonsson, *Some properties of 2-critically finite holomorphic maps of P^2* , Ergodic Theory Dynam. Systems 18 (1998), 171–187.
 - [Ka] A. Kameyama, *On Julia sets of postcritically finite branched coverings. I. Coding of Julia sets*, J. Math. Soc. Japan 55 (2003), 439–454.
 - [KH] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Cambridge Univ. Press, Cambridge, 1995.
 - [McM] C.T. McMullen, *Complex dynamics and renormalization.*, Princeton Univ. Press, Princeton, NJ, 1994.
 - [Me02] D. Meyer, *Quasisymmetric embedding of self similar surfaces and origami with rational maps*, Ann. Acad. Sci. Fenn. Math. 27 (2002), 461–484.
 - [Me09a] D. Meyer, *Dimension of elliptic harmonic measure of snowspheres*, Illinois J. Math. 53 (2009), 691–721.
 - [Me09b] D. Meyer, *Invariant Peano curves of expanding Thurston maps*, Preprint, 2009.
 - [Me09c] D. Meyer, *Expanding Thurston maps as quotients*, Preprint, 2009.
 - [Me10] D. Meyer, *Snowballs are quasiballs*, Trans. Amer. Math. Soc. 362 (2010), 1247–1300.
 - [Mi00] J. Milnor, *Local connectivity of Julia sets: expository lectures*, The Mandelbrot set, theme and variations, London Math. Soc. Lecture Note Ser., 274, Cambridge Univ. Press, Cambridge, 2000, 67–116.
 - [Mi04] J. Milnor, *Pasting together Julia sets: a worked out example of mating*, Experiment. Math. 13 (2004), 55–92.
 - [Mi06] J. Milnor, *On Lattès maps*, in: Dynamics on the Riemann Sphere, Eur. Math. Soc., Zürich, 2006, pp. 9–43.
 - [Mi] J. Milnor, *Dynamics in one complex variable*, Princeton Univ. Press, Princeton, NJ, 2006.
 - [Moi] E.E. Moise, *Geometric topology in dimensions 2 and 3*, Springer, New York, 1977.
 - [Moo] R.L. Moore, *Concerning upper semicontinuous collections of compacta*, Trans. Amer. Math. Soc. 27 (1925), 416–426.
 - [Ne] V. Nekrashevych, *Self-similar Groups*, Amer. Math. Soc., Providence, RI, 2005.
 - [Poi] A. Poirier, *On postcritically finite polynomials. Part two: Hubbard trees*, Stony Brook preprint no. 1993/7.
 - [Po] C. Pommerenke, *Boundary behaviour of conformal maps*, Springer, Berlin, 1992.
 - [Prz] F. Przytycki, *Hausdorff dimension of harmonic measure on the boundary of an attractive basin for a holomorphic map*, Invent. Math. 80 (1985), 161–179.
 - [Re] M. Rees, *Positive measure sets of ergodic rational maps*, Ann. Sci. École Norm. Sup. (4) 19 (1986), 383–407.

- [Ri] S. Rickman, *Quasiregular mappings*, Springer, Berlin, 1993.
- [Ru] W. Rudin, *Real and complex analysis*, 3rd ed., McGraw-Hill, New York, 1987.
- [TuV] P. Tukia and J. Väisälä, *Quasisymmetric embeddings of metric spaces*, *Ann. Acad. Sci. Fenn. Ser. A I Math.* 5 (1980), 97–114.
- [Vä] J. Väisälä, *Lectures on n -dimensional quasiconformal mappings*, Springer, Berlin-Heidelberg-New York, 1971.
- [Wa] P. Walters, *An introduction to ergodic theory*, Springer, New York-Berlin, 1982.
- [Yi] Q. Yin, *Lattès maps and combinatorial expansion*, in preparation.

INDEX

- Ahlfors condition, 178
- Ahlfors regular, 229
- bi-Lipschitz, 63
- branched covering map, 18
- cell, 21
 - decomposition, 21
 - of 2-sphere, 28
 - induced by Thurston map, 41
 - refinement, 24
- cell complex, 22
- cellular map, 25
- cellular Markov partition, 25
- chordal metric, 178
- closed equivalence relation, 127
- combinatorial expansion factor, 189
- combinatorially expanding, **97**, 125
- $\text{crit}(f)$, 18
- critical point, 18
- critical value, 18
- cycle of vertex, 30
- δ_0 , 52
- D_n , 54
- doubling, 178
- e -chain, 149
- e -connected, 149
- edge flower, 50
- edge path, 147
- \mathbf{E}^n , 42
- entropy
 - metric, 214
 - topological, 212
- ergodic, 213
- expanding, **19**, 44, 55, 56, 91, 201
 - combinatorially, **97**, 125
- expansion factor, 63
- factor of a dynamical system, 70
- flag, 29
- flower, 47
- generator, 214
- graph, 146
- Hausdorff convergence, 167
- Hausdorff distance, 167
- invariant measure, 212
- invariant set, 89
- isotopy, 20, 77
 - lift, 78
 - of Jordan curve, 84
- iterated monodromy group, 235
- joining opposite sides, **52**, 151
- Λ_0 , 189
- labeling, 99
 - compatible, 100
 - orientation-preserving, 100
- Lattès map, 9
- length of cycle, 30
- linearly locally connected (*LLC*), 183
- local degree, 18
- $m_{f,c}$, 60
- measure
 - invariant, 212
 - of maximal entropy, 215, 226
 - regular, 212
- metric entropy, 214
- mixing, 213
- Moore's Theorem, 127
- n -
 - cell, 42
 - edge, 42
 - edge flower, 50
 - flower, 47
 - skeleton, 22
 - tile, 42
 - vertex, 42
- order of cell, 42
- $\text{post}(f)$, 19
- postcritical point, 18
- postcritically-finite, 19
- quasiarc, 181
 - uniform, 181
- quasicircle, 178
 - uniform, 181

quasisymmetry, 178
 uniform, 180
 weak, 208

 refinement of cell decomposition, 24
 Rickman's rug, 121

 skeleton, 22
 snowflake equivalent, 63
 subdivision, 103
 subshift of finite type, 70

 Thurston equivalent, **20**, 81
 Thurston map, 19
 expanding, 19
 obstructed, 119
 topological entropy, 212
 topologically conjugate, **21**, 81
 two-tile subdivision rule, 103

 visual metric, **63**, 64
 \mathbf{V}^n , 42

 weak quasisymmetry, 208

 X_w^0, X_b^0 , 41
 \mathbf{X}^n , 42

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES,
LOS ANGELES, CA 90095, USA

E-mail address: mbonk@math.ucla.edu

UNIVERSITY OF HELSINKI, DEPARTMENT OF MATHEMATICS AND STATISTICS,
P.O. BOX 68, FI-00014 UNIVERSITY OF HELSINKI, FINLAND

E-mail address: dmeyermail@gmail.com